


На правах рукописи 

Храмеева Екатерина Евгеньевна

**АРХИТЕКТУРА ХРОМАТИНА И ЕЕ
РЕГУЛЯТОРНАЯ РОЛЬ В КЛЕТКАХ
ГОЛОВНОГО МОЗГА**

Специальность 1.5.8 —
«Математическая биология, биоинформатика»

Автореферат
диссертации на соискание учёной степени
доктора биологических наук

Москва — 2023

Работа выполнена в Автономной некоммерческой образовательной организации высшего образования "Сколковский институт науки и технологий".

Научный консультант:

Гельфанд Михаил Сергеевич, доктор биологических наук, профессор, Автономная некоммерческая образовательная организация высшего образования "Сколковский институт науки и технологий".

Официальные оппоненты:

Пономаренко Елена Александровна, доктор биологических наук, Федеральное государственное бюджетное научное учреждение "Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича".

Шайтан Алексей Константинович, доктор физико-математических наук, профессор РАН, член-корреспондент РАН, Биологический факультет Московского государственного университета имени М. В. Ломоносова.

Москалев Алексей Александрович, доктор биологических наук, профессор РАН, член-корреспондент РАН, Институт биологии Коми научного центра Уральского отделения Российской академии наук Федерального государственного бюджетного учреждения науки Федерального исследовательского центра «Коми научный центр Уральского отделения Российской академии наук».

Ведущая организация:

Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук».

Защита состоится 4 марта 2024 года в 15:00 часов на заседании диссертационного совета 24.1.101.01 на базе Федерального государственного бюджетного учреждения науки Института проблем передачи информации им. А.А. Харкевича Российской академии наук (ИППИ РАН) по адресу: г. Москва, Большой Каретный переулок, д.19, стр.1.

С диссертацией можно ознакомиться в библиотеке ИППИ РАН и на сайте ИППИ РАН: <http://iitp.ru/ru/dissertation/1650.htm>.

Автореферат разослан « »

2024 года.

Ученый секретарь
диссертационного совета
д.б.н.

Казенников Олег Васильевич

Общая характеристика работы

Актуальность темы. Геномы млекопитающих характеризуются сложной 3D архитектурой с множеством уровней организации, представляющих собой многослойную систему с определенной функциональностью. С развитием технологии захвата конформации хромосом (3C) было обнаружено, что эукариотические геномы организованы иерархически [1]. От крупного к мелкому масштабу, в ядре представлены хромосомные территории, компартменты хроматина, топологически ассоциированные домены (ТАДы), петли хроматина и нуклеосомы (рис. 1А). Индивидуальные хромосомы четко разделены в трехмерном пространстве ядра, с образованием хромосомных территорий - ядерных областей, преимущественно занятых различными интерфазными хромосомами [2]. Данные, полученные с помощью технологии высокопроизводительного захвата конформации хромосом (Hi-C, рис. 1Б) показали, что хромосомы дополнительно разделены на два компартмента. Компартмент А состоит из активных эпигенетических меток и активно транскрибируемых генов, обычно расположенных в центре ядра. Компартмент В состоит из репрессивных эпигенетических меток и неактивных генов, расположенных рядом с ядерной ламиной [3]. Считается, что формирование компартментов А и В обусловлено комбинацией факторов, включающей распределение активных и репрессивных меток хроматина [4].

Компартменты хроматина дополнительно разделены на ТАДы - геномные области с частыми взаимодействиями внутри них и высокой изоляцией от соседних ТАДов [5]. Предполагается, что ТАДы являются основными функциональными регуляторными доменами, модулирующими контакты между энхансерами и промоторами. Повышенная частота контактов внутри ТАДов опосредует физическое взаимодействие между парами энхансер-промотор, в то время как высокая изоляция на границах ТАДов ограничивает такие взаимодействия для энхансер-промоторных пар, расположенных в соседних ТАДах. Границы ТАДов очень стабильны между видами и обогащены фактором связывания СССТС (СТСФ) и когезином у млекопитающих. СТСФ и когезин образуют петли хроматина, которые закрепляют границы ТАДов, обеспечивая структурную основу для образования ТАДов. Считается, что петли хроматина формируются с помощью модели петлевой экструзии, основанной на взаимодействии СТСФ и комплекса когезина [6]. Физические взаимодействия между дистальными участками генома позволяют линейно удаленным друг от друга элементам, таким как промоторы и энхансеры, встретиться друг друга. Дальние контакты между энхансерами и промоторами могут регулироваться транскрипционными факторами, белком Mediator, РНК-полимеразой II и некодирующими РНК, которые дополнительно регулируют экспрессию генов [7].

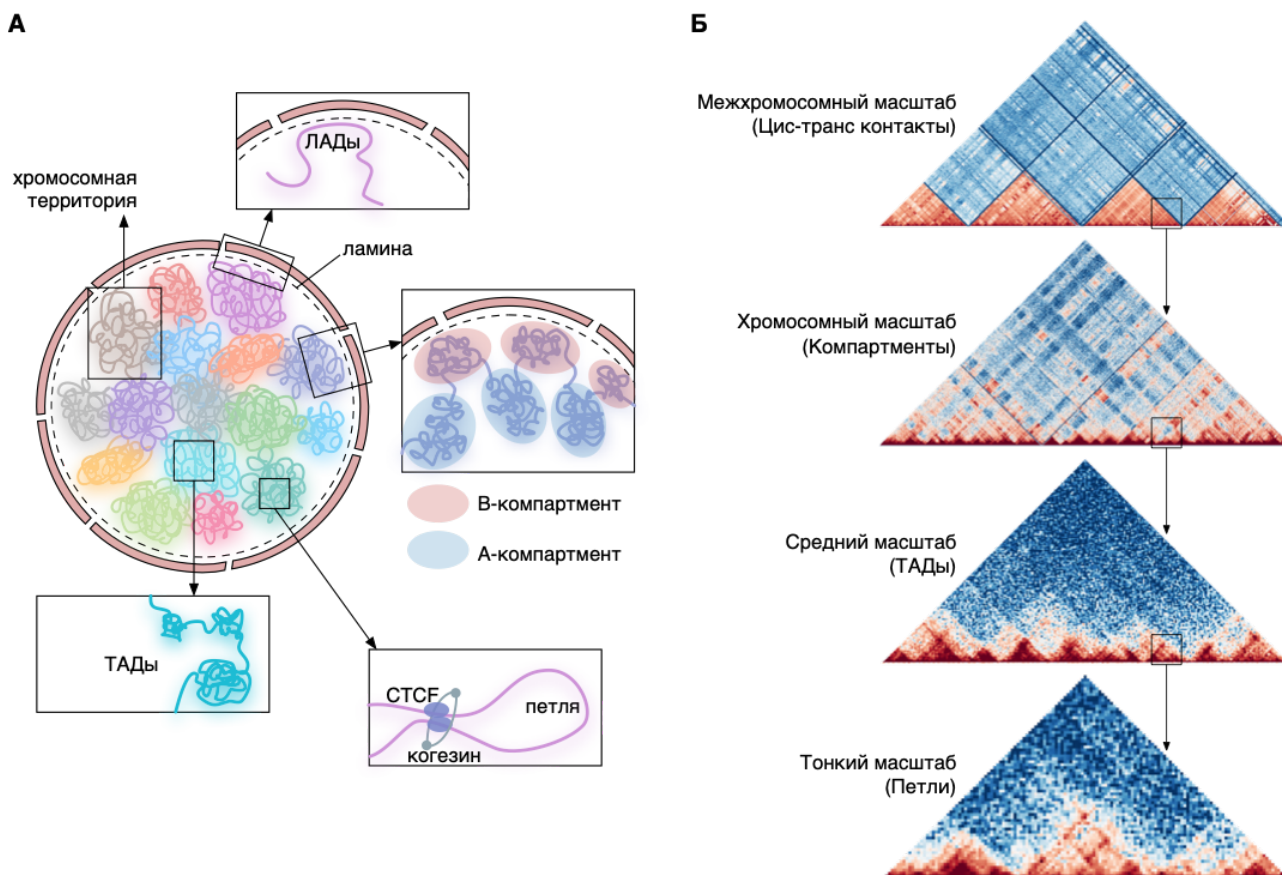


Рис. 1 — Общие принципы организации хроматина. (А) Схематичное изображение основных уровней организации хроматина в ядре. (Б) Уровни организации хроматина в данных Hi-C. Яркость пикселей на картах Hi-C пропорциональна частотам контактов.

Хотя трехмерная структура хроматина глобально стабильна, недавние исследования показывают, что отдельные гены часто переключаются между активными и неактивными компартментами во время развития. Кроме того, специфические взаимодействия внутри и вне ТАДов часто изменяются [8]. Наблюдаемые ассоциации между компартиментализацией хроматина, профилем ТАДов и транскрипционной активностью указывают на возможную причинно-следственную связь между ними. Так, сравнение карт Hi-C у плодовой мушки дрозофилы до и после подавления транскрипции показывает, что трехмерная структура хроматина регулирует экспрессию генов и, в свою очередь, также может быть модулирована генами, связанными с транскрипцией [9; 10]. Таким образом, сложное устройство архитектуры хроматина внутри ядра является важнейшим аспектом регуляции различных клеточных процессов и поддержания целостности генома.

Учитывая тесную связь архитектуры генома с экспрессией генов, изучение деталей трехмерной организации генома могло бы прояснить механизмы, лежащие в основе тканеспецифичности, а также продвинуть понимание патологических процессов, приводящих к заболеваниям. Таким образом, исследования, сосредоточенные на мозговой ткани, являются одними из самых востребованных в настоящее время ввиду недостаточного понимания сложно-

го процесса развития, функционирования мозга и его дегенеративных изменений. Мало работ исследовали разницу в архитектуре генома между нейронами и различными глиальными клетками [11; 12]. Однако они заложили основу исследований хроматина в головном мозге, продемонстрировав, что, хотя клетки мозга в целом соответствуют фундаментальным принципам трехмерной организации генома, они также демонстрируют уникальные особенности архитектуры хроматина.

В контексте этих исследований, представляется важным изучать молекулярную организацию мозга человека комплексно и анализировать организацию хроматина совместно с экспрессией генов, эпигенетическими модификациями, связыванием транскрипционных факторов, открытостью хроматина и метаболизмом клеток в целом. Сопоставление данных Hi-C с другими омиксными данными позволяет определять механизмы регуляции экспрессии генов, опосредованные трехмерной структурой хроматина. Архитектура хроматина организует контакты энхансеров и промоторов и таким образом регулирует уровни экспрессии генов, которые определяют количество структурных белков, ферментов и, опосредованно, синтезируемых ими метаболитов, что в комплексе и определяет фенотип и особенности функционирования клеток мозга человека.

Таким образом, несмотря на центральную роль трехмерной организации генома в регуляции функционирования человеческого мозга, исследования в этой области ограничены сложностью его анатомии и гетерогенностью его клеток. Тем не менее, с появлением высокопроизводительного секвенирования и усовершенствованных методов изучения пространственной организации генома теперь стало возможным исследовать такие сложные образцы. Несколько недавних работ начали проливать свет на ранее неизвестные аспекты архитектуры хроматина в различных типах клеток головного мозга, а также на роль организации хроматина в регуляции экспрессии генов. Однако для выяснения деталей организации хроматина в основных типах клеток головного мозга требуются всесторонние исследования в этой области. Результаты таких исследований могли бы предоставить ценную информацию об основных механизмах функционирования мозга как на клеточном, так и на органном уровнях.

Целью данной работы является установление роли трехмерной организации генома в регуляции функционирования человеческого мозга, с учетом гетерогенности его клеточного состава и в комплексе с другими уровнями молекулярной организации клеток: экспрессией генов, эпигенетическими модификациями, связыванием транскрипционных факторов, открытостью хроматина и метаболизмом клеток в целом.

Для достижения поставленной цели необходимо было решить следующие **задачи**:

1. Разработать методологию биоинформатического анализа данных Hi-C для образцов мозга человека: построить карты контактов, провести анализ хроматиновых компартов, ТADов и петель.

2. Решить проблему разной экспериментальной представленности геномных регионов в данных Hi-C.

3. Разработать методику выявления биологического сигнала в данных Hi-C с высоким уровнем технической вариабельности.

4. Оптимизировать метод аннотации ТADов на основе обогащения эпигенетическими метками.

5. Адаптировать и применить методы машинного обучения для исследования организации хроматина.

6. С помощью разработанных подходов к анализу данных Hi-C, описать особенности трехмерной организации генома в нейронах и других клетках мозга человека, а также установить функциональную роль дальних взаимодействий, которые опосредованы PcG белками и присутствуют только в нейронах.

7. Раскрыть общие принципы организации хроматина, такие как иерархичность ТADов, в том числе в одиночных клетках, роль ТADов в регуляции транскрипции, роль ядерной ламины, и др., чтобы поместить наблюдения для нейронов в контекст общих представлений об организации хроматина.

8. Разработать методы анализа данных RNA-seq, ATAC-seq, ChIP-seq и др. для решения нестандартных задач в контексте изучения мозга человека, таких как анализ биаллельной транскрипции, механизмы регуляции альтернативного сплайсинга (данные iCLIP), анализ некодирующих РНК, разделение сигнала от тканей человека и мыши в ксенотрансплантатах (PDX) для изучения онкозаболеваний, вычислительное восстановление видовой принадлежности отдельных клеток (данные snRNA-seq).

9. Разработать методы анализа липидного и метаболического состава мозга человека и с помощью них установить роль липидов и конкретных метаболитов в функционировании клеток головного мозга человека в норме и при заболеваниях, в частности при психических расстройствах.

10. Оценить роль неандертальских вариантов в качестве возможной причины липидных отличий, специфичных для европейской популяции, и оценить роль неандертальских вариантов в других популяциях, в частности в популяциях кетов и папуасов.

11. Разработать методы интеграции данных Hi-C с другими омиксными данными: RNA-seq, ATAC-seq, ChIP-seq и др., при помощи которых описать функционирование клеток мозга комплексно, на различных уровнях молекулярной организации, чтобы установить регуляторную роль деацетилазы SIRT6 и архитектурного белка YY1, участвующего в образовании хроматиновых петель.

12. Установить роль организации хроматина в регуляции человек-специфичной экспрессии генов в мозге человека.

Основные положения, выносимые на защиту:

1. Разработанная нами методика учета надежности контактов на основе сравнения экспериментальных повторностей позволяет выявлять биологический сигнал в данных Hi-C с высоким уровнем технической вариабельности. Суммарная частота контактов геномного региона в картах Hi-C является его биологическим свойством, а не технической погрешностью.

2. Добавление функциональной информации о хроматиновых метках и архитектурных белках к структурной информации о контактных частотах позволяет получать оптимизированную аннотацию топологических доменов.

3. ТАДы организованы иерархически в тканях, но не в одиночных клетках. Их границы плохо предсказываются по профилю связывания CTCF у дрозофилы, в отличие от человека, а значит в формирование ТАДов вносят вклад другие факторы, такие как ацетилирование гистонов. Позиции связывания белка Chromator и гистоновые модификации H3K4me3, H3K27ac являются наиболее информативными признаками для предсказания профиля ТАДов, как показали методы машинного обучения.

4. Разрушение ламины увеличивает плотность хроматина во фракции ТАДов, обогащенных активным хроматином, и приводит к отдалению хроматина от ядерной оболочки. На примере сперматоцитов мы показали, что инициация транскрипции коррелирует с отдалением промоторов от ламины и с локальной пространственной изоляцией прилегающих к ним областей. А активация транскрипции гена альфа-глобина коррелирует с усилением регуляции нескольких соседних генов домашнего хозяйства, таким образом влияя на структуру хроматина не только локально.

5. Хроматин имеет специфическую структуру в нейронах, по сравнению с другими клетками мозга человека: хромосомные территории более выражены, при одновременном драматическом снижении компартиментализации. ТАДы более выражены в нейронах по сравнению с другими клетками мозга, что объясняется повышенной концентрацией когезина в нейронах, а нейрональные петли значительно длиннее, что согласуется с обогащением загрузчика когезина NIPBL в нейронах. Дальние взаимодействия, опосредованные PcG белками, присутствуют только в нейронах, где выполняют функцию подавления развития транскрипционных факторов.

6. Построение подробной транскриптомной карты мозга человека в сравнении с другими приматами позволило установить, какие регионы у человека быстрее эволюционируют: кора больших полушарий, гипоталамус и серое и белое вещества мозжечка. Анализ с разрешением в одну клетку показал, что глиальные клетки эволюционировали быстрее нейронов на человеческой линии.

7. Разработанные нами методики анализа данных RNA-seq, ATAC-seq, ChIP-seq, iCLIP, DamID позволяют решать нестандартные задачи в контексте изучения мозга человека, например: анализ биаллельной транскрипции и механизмов регуляции альтернативного сплайсинга примере гена SF1 человека, который играет ключевую регуляторную роль в нейронах гипоталамуса; анализ некодирующих РНК, которые важны для функционирования мозга человека и, как мы показали, отличаются человек-специфичной экспрессией; разделение сигнала от тканей человека и мыши в ксенотрансплантатах для изучения онкозаболеваний.

8. С их помощью, а также с использованием разработанного нами метода анализа липидного и метаболического состава мозга человека, мы установили, что деацетилаза SIRT6 играет ключевую роль в регуляции митохондриальных процессов в головном мозге как напрямую, так и посредством образования комплекса с архитектурным белком YY1, участвующем в образовании хроматиновых петель.

9. В сравнении с другими видами млекопитающих, у человека липидные изменения локализованы преимущественно в мозге и вовлечены в метаболические пути, связанные с когнитивными заболеваниями, такими как болезнь Альцгеймера, Паркинсона, и другими нарушениями работы нервной системы. Причем концентрации липидов и экспрессия генов значительно изменялись от шимпанзе к человеку преимущественно в метаболических путях, связанных с катаболизмом липидов, у европейцев по сравнению с другими популяциями.

10. Анализ геномных последовательностей генов липидного катаболизма показал, что частота неандертальских вариантов выше в три раза в этих генах, причем только у европейцев (20,8%), по сравнению со средним по геному значением (6,1%). По этим генам шел положительный отбор у современных европейцев, но не у азиатов и африканцев.

11. В популяции кетов повторного события скрещивания с неандертальцами не происходило, а сохранение аллелей неандертальцев и денисовцев в геномах папуасов могло быть связано с адаптацией. Детальное исследование отличий в липидном и метаболитном составе головного мозга между современными популяциями продемонстрировало значимые различия у представителей китайской народности Хань, причем в основном в путях, связанных с липидным метаболизмом.

12. При аутизме, представляющем собой нарушение эволюционно новых механизмов работы мозга, наблюдаются существенные изменения метаболизма в префронтальной коре головного мозга, достигающие 15% от общего количества измеренных метаболитов. При ответе на антидепрессант флуоксетин, применяемый при депрессии, также происходят существенные изменения метаболизма в префронтальной коре, включая значимые изменения в количестве 106 липидов, сопровождающиеся незначительными изменениями в экспрессии генов.

13. Метаболический путь пуринов нарушается при развитии аутизма и при этом является эволюционно новым. Этот путь содержит фермент аденилосукцинатлиазу (ADSL), несущий аминокислотную замену, которая присутствует у современных людей, но отсутствует у неандертальцев. С использованием гуманизированной мышиной модели мы показали, что изменение пуринового метаболизма произошло у людей после их отделения от предка, общего с неандертальцами и денисовцами.

14. В целом, ускорение эволюции экспрессии на человеческой ветке связано с консервативными участками генома, накопившими человек-специфические отличия в последовательности. Эти участки обогащены энхансерами генов с человек-специфической экспрессией, и их регуляция может быть опосредована архитектурой хроматина.

Научная новизна. Одной из нерешенных задач в исследованиях головного мозга является соотнесение молекулярной структуры мозга с его функциональными свойствами. Особенно сложно это в случае головного мозга человека, который характеризуется с одной стороны высочайшей степенью сложности своей организации, а с другой стороны множественным спектром выполняемых когнитивных функций. Исследования, проведенные ранее на уровне экспрессии генов, позволили выявить множественные отличия в экспрессии, уникальные для человеческого мозга, что привело в итоге к открытию возможных механизмов, лежащих в основе когнитивной уникальности человека, например, более продолжительный синаптогенез и удлиненный период формирования синапсов в неокортексе у человека [13; 14]. Однако, в то же время, исследования, проводимые на уровне экспрессии генов, не позволили выявить функциональные особенности, связанные с теми многообразными функциями в когнитивных процессах, которые выполняют различные регионы головного мозга, в частности, различные зоны неокортекса. Несмотря на то, что исследования на уровне экспрессии, использующие такие методические подходы как секвенирование РНК, выделенной из единичной клетки или даже единичного ядра, и позволяющие проводить изучение организации мозга с высокой разрешающей способностью, могут пролить свет на имеющиеся различия в молекулярной организации между зонами неокортекса, становится очевидным, что глубокое понимание того, как функционирует мозг, невозможно без проведения всестороннего анализа молекулярной организации головного мозга. Важный уровень молекулярной организации мозга, которому до недавнего времени не было уделено внимания в работах, проводимых в данной области, - это трехмерная архитектура генома. В настоящей работе:

1. Разработан ряд новых подходов для анализа данных Hi-C, направленных на понижение технической вариабельности и решение других технических проблем в данных Hi-C, в том числе с использованием современных

методов машинного обучения, а также методика интеграции данных Hi-C с другими типами омиксных данных.

2. Применение разработанных подходов позволило впервые изучить трехмерную организацию хроматина в нейронах и других клетках региона ВА22р коры мозга человека и показать, что хроматин имеет специфическую структуру в нейронах, по сравнению с другими клетками мозга, вследствие чего многие биологические процессы, включая экструзию петель, компартиментализацию, дальние взаимодействия, вероятно, работают в нейронах по-другому, что способствует правильной экспрессии генов в клетках этого типа.

3. Впервые подробно описаны дальние взаимодействия в нейронах, которые опосредованы PcG белками и практически отсутствуют в других типах клеток, и установлена их функциональную роль, которая заключается в подавлении развитийных транскрипционных факторов.

4. Раскрыты функциональные свойства и механизм формирования ТАДов и показано, что ТАДы организованы иерархически в тканях, но не в одиночных клетках, и что их границы плохо предсказываются по профилю связывания CTCF у дрозофилы, в отличие от человека, а значит в формировании ТАДов могут вносить вклад другие факторы, такие как ацетилирование гистонов.

5. Впервые показана ключевая роль ядерной ламины в организации хроматина в целом и ТАДов в частности, и продемонстрирована важность ТАДов для активации транскрипции на примере сперматогенеза и альфа-глобинового локуса.

6. Разработаны методики анализа данных RNA-seq, ATAC-seq, ChIP-seq и др. для решения нестандартных задач в контексте изучения мозга человека, таких как анализ биаллельной транскрипции, механизмы регуляции альтернативного сплайсинга (данные iCLIP), анализ некодирующих РНК, разделение сигнала от тканей человека и мыши в ксенотрансплантатах (PDX) для изучения онкозаболеваний, вычислительное восстановление видовой принадлежности отдельных клеток (данные snRNA-seq).

7. Впервые разработан метод анализа липидного и метаболического состава мозга человека, применение которого показало, что деацетилаза SIRT6 играет ключевую роль в регуляции митохондриальных процессов в головном мозге через взаимодействие с архитектурным белком YY1, участвующим в образовании хроматиновых петель.

8. С помощью этой же методики впервые найдены: (а) липиды, специфические для человека по сравнению с другими видами, в том числе в головном мозге; (б) отличия в липидном и метаболическом составе головного мозга между популяциями современных людей; (в) существенные метаболические изменения при аутизме и при ответе на антидепрессант флуоксетин в коре головного мозга.

9. Впервые подробно описан метаболический путь биосинтеза пуринов и участвующий в нем фермент аденилосукцинаталиаза (ADSL), несущий аминокислотную замену, которая присутствует у современных людей, но отсутствует у неандертальцев, а также определено влияние этой замены на концентрации метаболитов из пути биосинтеза пуринов.

10. Разработана методика интеграции омиксных данных разных типов, с помощью которой впервые показано, что концентрации липидов и экспрессия генов значимо изменялись от шимпанзе к человеку в метаболических путях, связанных с катаболизмом липидов, у европейцев по сравнению с другими популяциями.

11. Впервые показано, что частота неандертальских вариантов выше среднего по геному значения в три раза в генах катаболизма липидов, причем только у европейцев. По этим генам шел положительный отбор у современных европейцев, но не у азиатов и африканцев, что указывает на возможную роль этих генов в адаптации к условиям окружающей среды доисторической Европы. Сохранение аллелей неандертальцев и денисовцев в геномах папуасов также могло быть связано с адаптацией.

12. Впервые установлено, что организация хроматина участвует в регуляции экспрессии генов в мозге человека, в особенности в участках генома, характеризующихся ускоренной эволюцией на человеческой ветке.

Научная и практическая значимость. Результаты данной работы приближают нас к пониманию того, как ДНК упакована внутри ядра клетки, и как особенности упаковки влияют на регуляцию генов, а значит помогают ответить на ключевые вопросы современной клеточной биологии и высокотехнологичной медицины. Решение этих вопросов необходимо для идентификации механизмов целого спектра социально значимых заболеваний человека, поскольку упаковка хромосом тесно связана с наследственностью и болезнями. Например, нарушение организации хромосом приводит к нарушениям развития и онкологическим заболеваниям: нейробластоме, лейкемии, медуллобластоме, глиоме, раку прямой кишки. Эти онкологические заболевания являются многофакторными, и у каждого пациента могут быть вызваны поломками в разных местах генома. Таким образом, изучение всего многообразия нарушений работы генов, связанных с изменениями в упаковке хромосом, необходимо при переходе к персональной медицине, для разработки новых высокотехнологичных (как и все современные методы лечения онкологических заболеваний) методов лечения и диагностики.

Получение карт конформации хромосом с высоким разрешением для образцов мозга человека само по себе имеет большую практическую значимость, поскольку закладывает основу для исследования причин патологий мозга, таких как психические расстройства, нейродегенеративные и онкологические заболевания. Предыдущие исследования указывали на важную роль архитектуры хромосом в заболеваниях мозга. Сопоставление данных о кон-

формации хромосом с экспериментами RNA-seq, ATAC-seq, ChIP-seq и другими позволяет интегрировать информацию о транскрипции, открытости и модификациях хроматина и изменениях конформации хромосом. Совместный анализ этих данных позволяет получать полную картину регуляторного ландшафта в нормальных клетках мозга человека и изучать возможные причины различных патологий, связанные с регуляторной ролью архитектуры хроматина.

Усовершенствование существующих и разработка новых подходов к анализу биологических данных по пространственной структуре хроматина, а также методов интеграции этих данных с разнородными данными по функциональным характеристикам генома позволяет изучать функции и механизмы формирования различных элементов организации хроматина (например, ТADов и хроматиновых петель) вычислительными методами без выполнения дорогостоящих экспериментов. Возможность сопоставления данных Hi-C с другими омиксными данными позволяет получать целостную картину изменений структурной и функциональной организации хромосом, связанных с тем или иным заболеванием, а значит приблизиться к пониманию механизмов заболеваний. Без детального понимания механизмов заболеваний невозможна разработка новых лекарственных препаратов и подходов к лечению заболеваний человека. Таким образом, понимание принципов и механизмов пространственной организации генома является одной из ключевых проблем современного здравоохранения.

Потенциал для практического применения данных Hi-C в медицине огромен: наличие Hi-C данных в нескольких временных точках позволяет проследить изменения в ходе заболевания в динамике. Например, проследить накопление геномных перестроек с течением онкологического заболевания, или изучить динамику изменений при нарушениях развития. С удешевлением секвенирования, которое активно происходит в настоящее время, получение таких данных Hi-C станет ещё более актуальным, в том числе для персонализированной медицины. А значит в будущем потребность в инструментах для анализа данных Hi-C в медицинских целях будет только возрастать.

Степень достоверности полученных результатов обеспечивается использованием современных экспериментальных и вычислительных методов с необходимым количеством валидационных экспериментов и контрольных анализов, а также подтверждается публикацией результатов в ведущих рецензируемых журналах. Сравнение с литературными источниками подтверждает, что наши результаты находятся в соответствии с результатами, полученными позднее другими авторами.

Апробация работы. Основные результаты работы докладывались на следующих международных и российских конференциях:

1. Международная конференция "18th annual international conference on Intelligent Systems for Molecular Biology (ISMB2010)". Бостон, США, 2010 г.,

устный доклад. Spatial proximity and similarity of functional states of genome domains. E. Khrameeva, A.A. Mironov, Ph. Khaitovich, and M.S. Gelfand.

2. Междисциплинарная школа-конференция "Информационные технологии и системы (ИТиС'10)". Геленджик, 2010 г., устный доклад. Functional similarity and chimeric transcripts in spatially close genome domains. E.E. Khrameeva, A.A. Mironov, M.S. Gelfand.

3. Международная конференция "24th International Mammalian Genome Conference". Крит, Греция, 2010 г., устный доклад. Association between spatial proximity and functional similarity in human genome. Ekaterina Khrameeva, Andrey Mironov, Mikhail Gelfand.

4. Международная конференция "Albany 2011: The 17th Conversation". Олбани, США, 2011 г. The Impact of Interchromosomal Associations on the Functional State of the Human Genome. Ekaterina E. Khrameeva, Andrey A. Mironov, Mikhail S. Gelfand.

5. Международная конференция "Московская конференция по вычислительной молекулярной биологии (МССМВ'11)". Москва, 2011 г., устный доклад. Interlaboratory and interplatform comparison of 117 mRNA and genome sequencing experiments. Ekaterina E. Khrameeva, Mikhail S. Gelfand.

6. Междисциплинарная школа-конференция "Информационные технологии и системы (ИТиС'11)". Геленджик, 2011 г., устный доклад. Comparison of 117 mRNA and genome sequencing experiments between laboratories and platforms. Ekaterina E. Khrameeva, Mikhail S. Gelfand.

7. Международная конференция "Cold Spring Harbor Asia. Bioinformatics of Human and Animal Genomics". Суджоу, Китай, 2011 г., устный доклад. Spatial proximity and similarity of the epigenetic state of genome domains. E. Khrameeva, A.A. Mironov, P. Khaitovich, and M.S. Gelfand

8. Международная конференция "RECOMB-seq 2012". Барселона, Испания, 2012 г. Biases in read coverage demonstrated by interlaboratory and interplatform comparison of 117 mRNA and genome sequencing experiments. Khrameeva EE, Gelfand MS.

9. Междисциплинарная школа-конференция "Информационные технологии и системы (ИТиС'12)". Петрозаводск, 2012 г., устный доклад. Regulation of alternative splicing by hnRNPL protein. Ekaterina E. Khrameeva.

10. Международная конференция "Chromosomes, Stem Cells and Disease". Барселона, Испания, 2012 г., постерный доклад. Spatial proximity and similarity of the epigenetic state of genome domain. Ekaterina E. Khrameeva.

11. Международная конференция "Московская конференция по вычислительной молекулярной биологии (МССМВ'13)". Москва, 2013 г., устный доклад. Association between changes in lipid catabolism and increased gene flow from Neanderthals to Europeans. Khrameeva E., Gelfand M., Khaitovich P.

12. Междисциплинарная школа-конференция "Информационные технологии и системы (ИТиС'13)". Калининград, 2013 г., устный доклад. Increased

similarity between Neanderthal and European genomes is associated with changes in lipid catabolism. E.E. Khrameeva, P. Khaitovich.

13. Междисциплинарная школа-конференция "Информационные технологии и системы (ИТиС'14)". Нижний Новгород, 2014 г., постерный доклад. Анализ структуры хроматина в двух клеточных линиях дрозофилы. E.E. Храмеева.

14. Международная конференция "Albany 2015: The 19th Conversation". Олбани, США, 2015 г., устный доклад. Active chromatin regions are sufficient to define borders of topologically associated domains in *D. melanogaster* interphase chromosomes. Khrameeva EE, Ulyanov SV, Gavrilov AA, Shevelyov YY, Gelfand MS, Razin SV.

15. Международная конференция "Московская конференция по вычислительной молекулярной биологии (МССМБ'15)". Москва, 2015 г., устный доклад. Active chromatin regions are sufficient to define borders of topologically associated domains in *D. melanogaster* interphase chromosomes. Khrameeva EE, Ulyanov SV, Gavrilov AA, Shevelyov YY, Razin SV.

16. Междисциплинарная школа-конференция "Информационные технологии и системы (ИТиС'15)". Сочи, 2015 г., устный доклад. Взаимосвязь хромосомных перестроек и пространственной структуры хроматина в дрожжах. Храмеева E.E., Фуденберг Д., Гельфанд М.С., Мирный Л.

17. Междисциплинарная школа-конференция "Информационные технологии и системы (ИТиС'16)". Репино, 2016 г., устный доклад. Молекулярные маркеры продолжительности жизни человека. E. Храмеева, И. Курочкин.

18. Всероссийская конференция "V Съезд биохимиков России". Дагомыс, 2016 г., устный доклад. Геномика аутизма. E. Храмеева, В. Степанова.

19. Международная конференция "шпез". Новосибирск, 2017 г., устный доклад. Гены катаболизма липидов обогащены неандертальскими вариантами у современных европейцев. E. Храмеева.

20. Международная конференция "Society for Molecular Biology and Evolution Meeting (SMBE'17)". Остин, США, 2017 г., постерный доклад. Evolution of lipid concentrations in mammalian tissues. E. Khrameeva, I. Kurochkin, P. Khaitovich.

21. Международная конференция "Московская конференция по вычислительной молекулярной биологии (МССМБ'17)". Москва, 2017 г., устный доклад. Factors shaping structural organization of chromatin. E. Khrameeva.

22. Международная конференция "Bioinformatics: from Algorithms to Applications (BiATA'17)". Санкт-Петербург, 2017 г., устный доклад. Studying chromatin 3D structure in eukaryotes. E. Khrameeva.

23. Международная конференция "Клиническая протеомика. Постгеномная медицина". Москва, 2017 г., устный доклад. Metabolic changes in autism and their evolution. E. Khrameeva.

24. Международная конференция "Московская конференция по вычислительной молекулярной биологии (МССМБ'19)". Москва, 2019 г., устный доклад. Transcriptome map of the human, chimpanzee, bonobo, and macaque brains at the single-cell resolution. E. Khrameeva, I. Kurochkin, P. Khaitovich.

25. Всероссийская конференция с международным участием "Школа по регенеративной медицине". Москва, 2021 г., устный доклад. Single-cell resolution transcriptomics: applications for the human brain research. E. Khrameeva.

26. Международная конференция "Bioinformatics of Genome Regulation and Structure/Systems Biology (BGRS/SB-2022)". Новосибирск, 2022 г., устный доклад. Maps of chromatin conformation in the cerebral cortex. Cherkasov A., Bazarevich M., Pletenev I., Efimova O., Kononkova A., Khrameeva E.

27. Всероссийская конференция с международным участием "Московская конференция по вычислительной молекулярной биологии (МССМБ'23)". Москва, 2023 г., устный доклад. Архитектура хроматина и опосредованные ей регуляторные механизмы в различных организмах и модельных системах. Е. Храмеева.

Личный вклад. В методических работах [Kobets и др., 2023; Smirnov и др., 2021; Rozenwald и др., 2020] автор диссертации предложила дизайн исследования и принимала прямое участие в разработке алгоритма, его тестировании, интерпретации результатов и документировании программной реализации. В работе [Pervouchine и др., 2012] автор диссертации принимала непосредственное участие в разработке алгоритма, написании программного кода, его тестировании и интерпретации результатов. В работах [Khrameeva и др., 2020; Kurochkin и др., 2019; Ulianov и др., 2019; Khrameeva и др., 2018; Vozek и др., 2017; Zhang и др., 2016; Ulianov и др., 2016; Khrameeva и др., 2016; Zhenilo и др., 2015; Yang и др., 2014; Khrameeva и др., 2014; Rossbach и др., 2014; Rösel-Hillgärtner и др., 2013; Khrameeva и Gelfand, 2012; Khrameeva и др., 2012] автором диссертации лично выполнен биоинформатический анализ экспериментальных данных, статистический анализ и интерпретация результатов анализа. В работах [Samborskaia и др., 2020; Eremenko и др., 2021; Стекольщикова и др., 2021; Akkuratov и др., 2018] автором диссертации поставлена задача, координирован процесс исследований и выполнена интерпретация полученных результатов. В работах [Smirnov и др., 2023; Shcherbinina и др., 2022; Plyin и др., 2022; Tkachev и др., 2021; Stein и др. 2021; Yegodayev и др., 2020] автор диссертации разработала общую схему анализа данных и принимала прямое участие в разработке и применении методик анализа омиксных данных различных типов. В работах [Stepanova и др., 2021; Mazor и др., 2021; Badarni и др., 2021; Ulianov и др., 2021] автор диссертации принимала непосредственное участие в разработке стратегии анализа данных, координации процесса исследований и интерпретации результатов, а также участвовала в выполнении биоинформатических частей исследований [Tkachev и др., 2019; Luzhin и др., 2019; Ulianov и др., 2017; Gavrillov и др., 2016; Flegontov и др.,

2016; Kaufmann и др., 2015]. Таким образом, автор принимала непосредственное участие в выборе стратегий исследований, обсуждении экспериментальной работы, планировании анализа данных и его непосредственном выполнении, интерпретации получаемых результатов, их обобщении, написании и редактировании текстов публикаций. В 9 статьях по теме диссертации автор выступает в качестве первого автора, и в 9 в качестве автора, ответственного за переписку. Все основные результаты анализа данных, описанные в диссертационной работе, получены автором лично или под ее непосредственным руководством.

Работа по теме диссертации была поддержана грантами РФФИ (21-74-10102, 19-74-00112) и РФФИ (21-34-70051, 20-34-70077), в которых автор являлась руководителем. Под руководством автора диссертации в рамках темы данной работы подготовлены и защищены 19 выпускных квалификационных работ специалистов и магистров. Работа по теме диссертации проводилась в сотрудничестве с Институтом проблем передачи информации РАН, Институтом биологии гена РАН, Институтом молекулярной генетики РАН, Институтом молекулярной и клеточной биологии Сибирского отделения РАН, Новосибирским государственным университетом, Биологическим факультетом, Физическим факультетом и Факультетом биоинженерии и биоинформатики МГУ имени М.В. Ломоносова, Национальным исследовательским центром "Курчатовский институт Центром "Биоинженерия" РАН, Университетом Гиссена (Германия), Институтом исследований сердца и легких Макса Планка (Германия), Массачусетским технологическим институтом (США), Университетом Людвиг-Максимилиана (Германия), Калифорнийским университетом (США), Университетом Остравы (Чехия), Детской больницей Лос-Анджелеса (США), Институтом психиатрии Макса Планка (Германия), Институтом эволюционной антропологии Макса Планка (Германия), Институтом биологии старения Макса Планка (Германия), Партнерским институтом вычислительной биологии (Китай), Университетом Бен-Гуриона в Негеве (Израиль).

Публикации. Основные результаты по теме диссертации изложены в 39 статьях в рецензируемых журналах, индексируемых Scopus и/или Web of Science.

Объем и структура работы. Диссертация состоит из введения, трех глав и заключения. Полный объем диссертации составляет **209** страниц текста с **46** рисунками и **3** таблицами. Список литературы содержит **232** наименования.

Содержание работы

Введение

Во введении обосновывается актуальность исследований, проводимых в рамках данной диссертационной работы, формулируется цель, ставятся задачи работы, перечисляются основные положения, выносимые на защиту, сформулированы научная новизна и практическая значимость представляемой работы, обсуждаются степень достоверности, апробация и личный вклад автора, приводится количество публикаций.

Обзор литературы

Первая глава посвящена обзору литературы по изучаемой проблеме. Обсуждаются общие принципы организации хроматина, от крупного к мелкому масштабу: хромосомные территории, компартменты хроматина, топологически ассоциированные домены (ТАДы), хроматиновые петли и другие уровни организации хроматина. Далее, приводится обзор современного состояния исследований об архитектуре хроматина в мозге человека. Обсуждаются особенности компартментализации, ТАДы и петли, а также особые уровни организации хроматина в нейронах и глиальных клетках. Рассматривается литература о регуляторной роли организации хроматина в мозге человека. Кроме того, представлены существующие методы изучения архитектуры хроматина, основанные на микроскопии, протоколах семейства 3С, и другие омиксные методы.

Материалы и методы

Вторая глава посвящена материалам и методам диссертационной работы. Сначала кратко представлены стандартные методы обработки данных Hi-C, а далее приводится подробное описание разработанных нами методов анализа данных Hi-C.

Суммарная частота контактов геномного региона является его биологическим свойством, а не технической погрешностью.

В одном из наших ранних исследований мы показали, что различные методы высокопроизводительного секвенирования подвержены технической вариабельности [Khrameeva и Gelfand, 2012]. Метод Hi-C не является исключением. Одним из недостатков анализа и интерпретации данных Hi-C является наличие систематических погрешностей, таких как различная доступ-

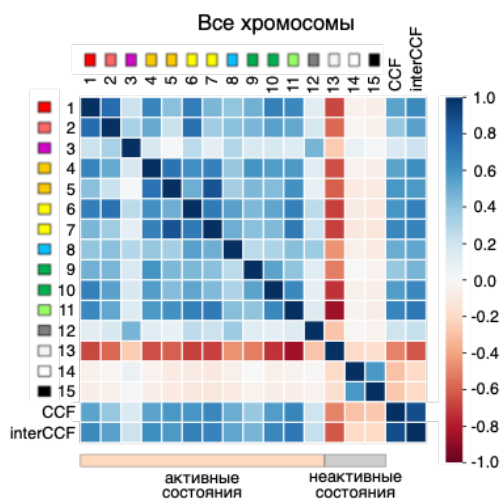


Рис. 2 — Матрица корреляций для состояний хроматина 1-15, внутривхромосомной CCF и межхромосомной CCF (interCCF) в геноме человека. Состояния 1-12 соответствуют активному хроматину, а состояния 13-15 - гетерохроматину. Цветовая шкала показывает коэффициенты корреляции Пирсона. Адаптировано из [Samborskaia и др., 2020].

ность для ферментов, амплификации и картируемости участков ДНК, что приводит к различной представленности геномных участков в данных Hi-C. Итеративная коррекция - один из наиболее популярных методов, разработанных для устранения этих систематических искажений. Она выравнивает экспериментальную представленность между различными участками генома, аппроксимируемую суммарной частотой контактов, далее сокращенно обозначаемой как CCF (Cumulative Contact Frequency). Однако различия в экспериментальной представленности могут быть объяснены не только техническими, но и биологическими факторами, такими как открытость хроматина, которая сильно варьируется между различными состояниями хроматина.

В работе [Samborskaia и др., 2020] нам удалось показать, что CCF положительно коррелирует с активной транскрипцией (рис. 2). Гетерохроматин, напротив, оказался отрицательно скоррелирован со всеми активными состояниями и с CCF (рис. 2). CCF также оказалась связана со структурой компартментов, поскольку компартмент А имеет более высокие уровни CCF и экспрессии генов, чем компартмент В [Samborskaia и др., 2020]. Примечательно, что указанные наблюдения справедливы для широкого спектра видов, включая человека, мышь и дрозофилу. Таким образом, глубокое изучение взаимосвязи между CCF и активностью хроматина помогло нам выбрать лучшую процедуру нормализации для последующего анализа данных Hi-C в мозге человека.

Выявление биологического сигнала в данных с высоким уровнем технической вариабельности.

Другой нерешенной технической проблемой в Hi-C данных является высокий уровень технической вариабельности, так называемый батч-эффект, который может приводить к низкому соответствию между экспериментальными повторностями. Поскольку для анализа архитектуры хроматина в мозге человека мы используем большое количество карт Hi-C, полученных в рамках

разных батчей (т.к. в рамках одного экспериментального батча Hi-C возможно получить не более 6 карт Hi-C), требовалось решить проблему батч-эффекта в данных. Для этого мы разработали новый подход - HiConfidence [Kobets и др., 2023]. Подход HiConfidence основан на учете «надежности» контактов между участками генома, на основе сравнения данных для технических повторностей одного и того же эксперимента. Будем учитывать только те контакты, которые имеют одинаковый биологический эффект. Вклад каждого пикселя на карте Hi-C в усредненную плотность рассчитывается с учетом его дисперсии (рис. 3).

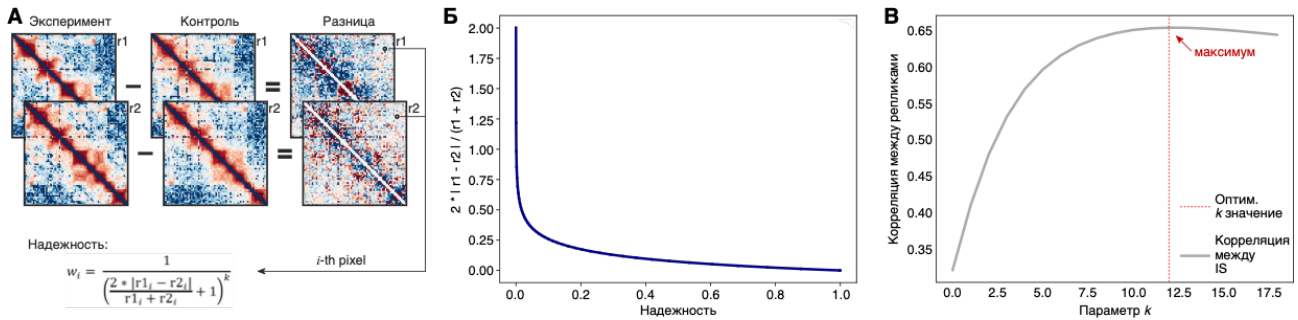


Рис. 3 — Описание алгоритма HiConfidence. (А) Ключевые этапы процедуры HiConfidence. Сперва разница между картами эксперимента и контроля Hi-C рассчитывается для каждой биологической повторности (r_1 и r_2) отдельно. Далее оценивается достоверность контакта для каждого пикселя полученных карт. (Б) Значение достоверности обратно пропорционально разнице между биологическими репликами, деленной на их среднее значение (в степени k). (В) Параметр k может быть оптимизирован путем максимизации коэффициента корреляции Пирсона между дифференциальными профилями инсуляции (IS) биологических повторностей. Адаптировано из [Kobets и др., 2023].

Поскольку данные Hi-C для мозга человека ещё не были доступны, мы протестировали подход HiConfidence на трех наборах данных Hi-C для клеток *D. melanogaster* S2 (рис. 4А). К нашему удивлению, кластеризация итеративно скорректированных карт Hi-C продемонстрировала группировку по экспериментальным батчам вместо биологических условий (рис. 4Б). Это противоречило нашей гипотезе о том, что все контрольные реплики должны образовывать отдельный кластер необработанных клеток. Реплики, представляющие собой деплецию деацетилазы гистонов (HDAC-dep) и ингибирование деацетилаз гистонов агентом TSA (HDAC-inh), должны были быть сгруппированы в отдельный кластер, соответствующий повышенному ацетилированию гистонов. А экспериментальные повторности ингибирования гистонацетилазы куркумином (HAT-inh) должны были образовать еще один отчетливый кластер, соответствующий пониженному ацетилированию гистонов. Поскольку такая группировка не наблюдалась, мы пришли к выводу, что анализируемые нами наборы данных Hi-C могут содержать в себе нежелательные технические эффекты.

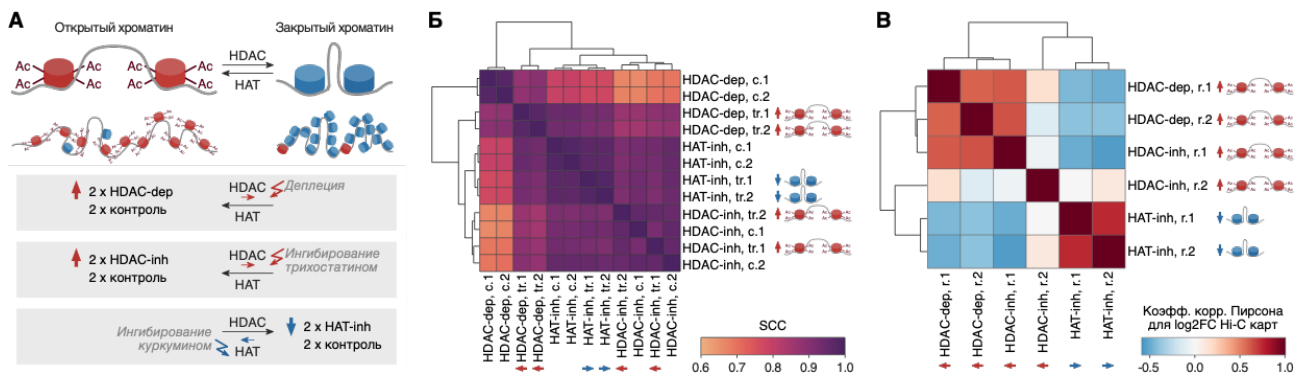


Рис. 4 — Технические проблемы мешают прямому сравнению наборов данных Hi-C. (А) Дизайн экспериментов Hi-C, используемых для оценки эффективности подхода HiConfidence. Красными стрелками показана деплеция деацетилазы гистонов (HDAC-dep) и ингибирование деацетилазы гистонов трихостатином А (HDAC-inh), повышающее уровень ацетилирования. Синими стрелками отмечено ингибирование гистоновых ацетилтрансфераз куркумином (HAT-inh), снижающее уровень ацетилирования. (Б) Корреляция карт Hi-C для обработанных (tr.) и контрольных (с.) повторностей, оцененная с использованием коэффициента корреляции с поправкой на ожидаемое (SCC). (В) Корреляция дифференциальных профилей инсуляции (IS) для карт \log_2FC Hi-C, представляющих собой разницу между обработанными репликами (r.) и соответствующими контрольными. Адаптировано из [Kobets и др., 2023].

Группировка по экспериментальным батчам вместо биологических условий указывает на то, что на экспериментальные повторности и контрольные реплики в составе одного экспериментального батча может влиять один и тот же технический фактор. Следовательно, анализ значений \log_2 -кратного изменения (\log_2FC) между обработанными и контрольными картами Hi-C внутри батча вместо отдельных карт Hi-C может уменьшить влияние технических факторов, аналогично процедуре сокращения дробей в школьной математике (рис. 4В). Эта кластеризация лучше соответствовала ожидаемой группировке образцов по биологическим условиям, за исключением одной повторности эксперимента HDAC-inh (r.2), которая продемонстрировала неожиданную кластеризацию и, по всей видимости, наиболее существенно пострадала от нежелательного технического шума.

После применения алгоритма HiConfidence к этому набору данных, карты \log_2FC Hi-C, представляющие собой разницу между обработанными повторностями и соответствующими контролями, объединились в кластеры, соответствующие биологическим условиям, в полном соответствии с ожидаемой группировкой.

Далее, мы провели сравнение нашего подхода с существующим методом нормализации multiHiScCompare на тех же данных, а также применили их к одному и тому же внешнему набору данных [15], в котором изучалось изменение хроматина на ранних этапах развития. В отличие от HiConfidence, применение multiHiScCompare существенно не улучшило согласованность между репликами в обоих наборах данных. Таким образом, подход HiConfidence является универсальным, может быть успешно применен к анализу данных

Hi-C в нескольких (более, чем двух) временных точках, и в дальнейшем может использоваться для анализа других данных Hi-C, в том числе для клеток мозга человека.

Метод аннотации ТАДов на основе обогащения эпигенетическими метками

Другой важной с методической точки зрения задачей является аккуратная разметка ТАДов в картах Hi-C. Существует множество инструментов для разметки ТАДов в картах Hi-C, однако все они имеют один или несколько параметров, от которых зависит получающаяся в результате разметка ТАДов. Проблема подбора параметров является чрезвычайно важной, т.к. параметры приходится подбирать вручную, что исключает повторяемость научных исследований: нет общепринятых правил подбора этих параметров, и каждая научная группа использует в своих исследованиях разные. Поэтому мы разработали универсальный и полностью автоматизированный подход к подбору параметров для алгоритмов разметки ТАДов - OptimalTAD.

Подход OptimalTAD основан на использовании комбинации Hi-C и эпигенетических данных (ChIP-seq, ChIP-chip) и поэтому наиболее подходит для исследования организмов, у которых наблюдается высокая корреляция между хромосомными контактами и эпигенетическими профилями. Предполагая, что наиболее выраженное обогащение соответствует лучшей аннотации границ ТАДов, наш алгоритм оптимизирует параметры поиска ТАДов, максимизируя разницу в уровнях меток хроматина между ТАДами и их границами. Используя этот алгоритм, мы аннотировали ТАДы в нескольких общедоступных наборах данных Hi-C и определили набор эпигенетических меток, которые лучше всего подходят для поиска ТАДов. Алгоритм использует программу Armatus в качестве базовой для предсказания набора ТАДов и оптимизирует ее внутренний параметр γ , определяющий средний размер предсказанных доменов (рис. 5).

Чтобы проиллюстрировать универсальность optimalTAD, мы протестировали optimalTAD на данных Hi-C и ChIP-seq из нашей работы [Ulianov и др., 2019] в контрольных клетках S2 *D. melanogaster* и при нокдауне ламина Dm0 (Lam-KD), а также на наборе данных Ing-Simmons [17], включающем карту Hi-C Tollrm9/rm10 и две реплики H3K27ac ChIP-seq. Кроме того, мы сравнили optimalTAD с несколькими ранее опубликованными алгоритмами поиска ТАДов. Среди всех алгоритмов, optimalTAD показал наибольшее значение Δh , превзойдя другие алгоритмы более чем в 1,7 раза. Соответственно, кривая ацетилирования, предсказанная с использованием optimalTAD, показала наиболее резкое различие в уровнях ацетилирования между ТАДами и межТАДами по сравнению с другими алгоритмами.



Рис. 5 — Общая схема работы алгоритма optimalTAD. Адаптировано из [16].

Таким образом, optimalTAD производит оптимизированную идентификацию топологических доменов и может успешно применяться к данным Hi-C, полученным из различных источников и для различных организмов, в том числе и для клеток мозга человека. В то время как другие алгоритмы поиска ТАДов ограничены использованием только структурной информации о контактных частотах и предсказанием границ на основе заранее определенного размера ТАДов, optimalTAD, кроме того, использует функциональную информацию о хроматиновых метках и архитектурных белках для аннотации ТАДов как структурных и функциональных единиц организации хроматина.

Применение методов машинного обучения для изучения особенностей организации хроматина

Мы применяли методы машинного обучения для определения факторов, отвечающих за формирование ТАДов у дрозофилы, ещё в 2016 году в работе [Ulianov и др., 2016]: использовали модель логистической регрессии для решения задачи классификации (рис. 6). Вход модели представлял собой набор сигналов ChIP-chip для области генома, а выход - бинарное значение, указывающее, была ли область расположена на границе или внутри ТАДа. Модель, основанная на распределениях H3K4me3 и тотальной РНК, предсказывала намного лучше, чем модель, основанная на распределениях dCTCF и Su(Hw) ($AUC = 0,79$ и $0,63$ соответственно). Причем использование dCTCF и Su(Hw) вместе с активными метками хроматина не улучшало качество предсказания ($AUC = 0,79$). Поскольку профили большинства хроматиновых меток сильно скоррелированы, добавление признаков, связанных с активным хроматином, также существенно не улучшало качество предсказания (рис. 6). Таким образом, эти результаты позволяют предположить, что модификации гистонов, характерные для активного хроматина и/или актив-

ной транскрипции, а не связывание белков-инсуляторов dCTCF и Su(Hw), являются основной характерной чертой границ ТАДов и областей между ТАДами в клетках дрозофилы.

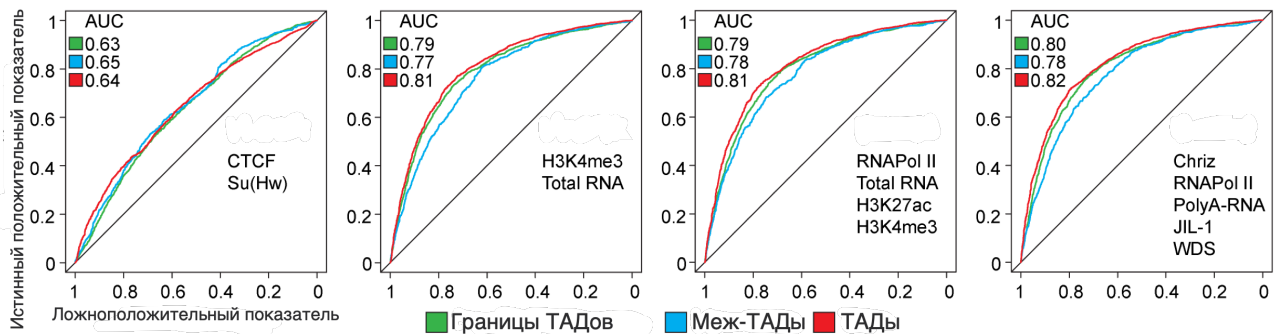


Рис. 6 — Предсказание ТАДов, межТАДов и границ ТАДов в клетках S2 дрозофилы с использованием моделей логистической регрессии, основанных на геномном распределении активных меток хроматина или архитектурных белков dCTCF и Su(Hw). Показаны кривые ROC и значения AUC (площадь под кривой). Адаптировано из [Ulianov и др., 2016].

В работе [Rozenwald и др., 2020], мы применили методы машинного обучения для предсказания паттернов сворачивания ДНК у дрозофилы на основе одномерных эпигенетических данных (хроматиновых меток) в трех клеточных линиях. Цель этого анализа состояла в том, чтобы предсказать профиль ТАДов по эпигенетическим маркерам и определить, какие из них являются наиболее значимыми для предсказания профиля ТАДов. Поэтому мы использовали модели линейной регрессии с четырьмя типами регуляризации, регрессионные модели с градиентным бустингом и рекуррентные нейронные сети (RNN) в качестве инструментов для изучения характеристик сворачивания хроматина, связанных с ТАДами, на основе данных ChIP-seq и ChIP-chip. Архитектура RNN с двунаправленной долгосрочной краткосрочной памятью (biLSTM) показала наилучшие результаты и выявила биологически значимые маркеры.

Распределение белка Chriz (Chromator) и гистоновые модификации H3K4me3, H3K27ac оказались наиболее информативными признаками для предсказания ТАДов (рис. 7), что указывает на потенциальную важность этих факторов для формирования ТАДов. Разработанный нами подход может быть адаптирован к любому аналогичному набору биологических данных о характеристиках хроматина в любом организме и типе клеток, в том числе для клеток мозга человека.

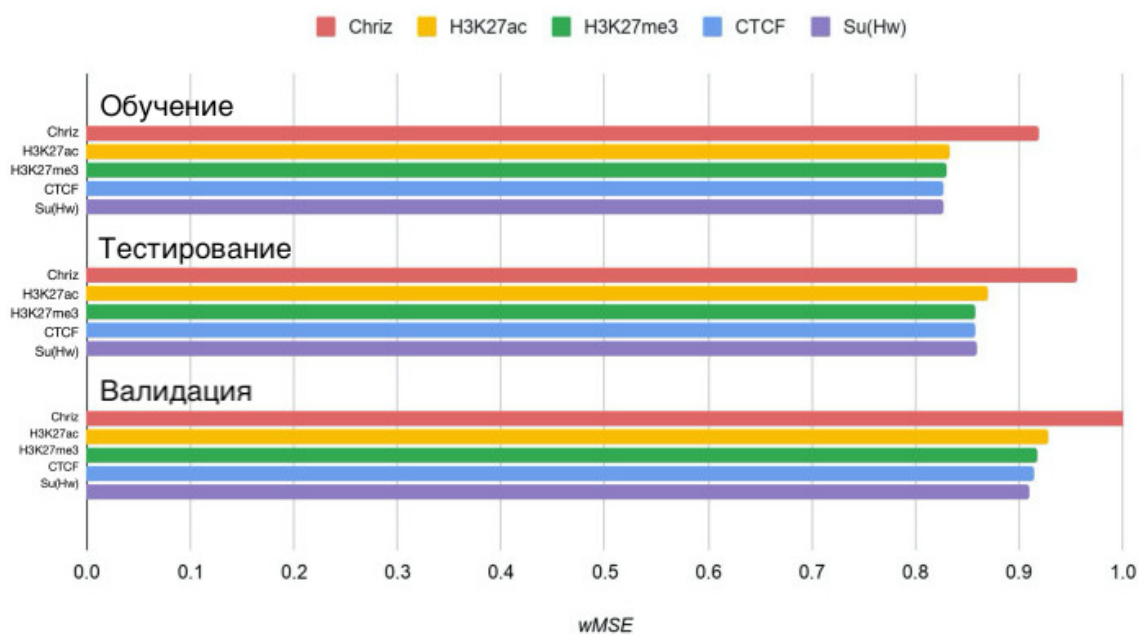


Рис. 7 — Взвешенная среднеквадратичная ошибка (wMSE) с использованием четырех из пяти хроматиновых маркеров в качестве входных данных biLSTM RNN. Каждый цвет соответствует признаку, который был исключен из входных данных. Обратите внимание, что на модель больше всего влияет исключение фактора Chriz. Адаптировано из [Rozenwald и др., 2020].

Разработка методов интеграции данных Hi-C с другими омиксными данными

Получение целостной картины структурной и функциональной организации хромосом невозможно без сопоставления данных Hi-C с другими омиксными данными: RNA-seq, ATAC-seq, ChIP-seq и другими.

Чтобы решить задачу ассоциации данных Hi-C с омиксными данными разных типов, мы разработали следующий подход: посчитали среднее содержание эпигенетического сигнала (высоты пика или среднего покрытия) в бинах Hi-C карты вокруг каждой границы ТАДов или хроматиновых петель. Полученные значения были дополнительно Z-трансформированы для сопоставимости распределений, полученных для различных экспериментов. В результате, мы построили средний профиль значений RNA-seq, ChIP-seq, ATAC-seq вокруг границ различных элементов укладки хроматина (ТАДов, хроматиновых петель и др.). Этот подход позволяет выявлять дифференциальное связывание эпигенетических факторов вокруг границ ТАДов, хроматиновых петель или других элементов организации хроматина.

Поскольку данные Hi-C для мозга человека ещё не были доступны, мы отработывали эту методику в работах [Ulianov и др., 2019; Pyin и др., 2022] на данных дрозофилы, геном которой в 10 раз меньше генома человека, что значительно уменьшает время расчетов. В первой работе [Ulianov и др., 2019] мы интегрировали данные Hi-C с данными RNA-seq и ChIP-seq как описано выше, чтобы изучить, как ядерная ламина влияет на глобальную архитектуру

туру хроматина. Благодаря разработанной методике анализа, нам удалось показать, что разрушение ламины в клетках дрозофилы приводит к уплотнению хроматина и его отдалению от ядерной оболочки. Это увеличивает плотность хроматина во фракции ГАДов, обогащенных активным хроматином, и усиливает взаимодействия между активным и неактивным хроматином. Стоит заметить, что позднее мы ещё возвращались к изучению влияния ламины на архитектуру хроматина в головном мозге, поскольку белок SIRT6, регуляторную роль которого в мозге мы исследовали в работах [Stein и др., 2021; Smirnov и др., 2023] (см. ниже в разделе Результаты и обсуждение), активируется белком ламины LaminA, обеспечивающим локализацию SIRT6 в хроматине при обнаружении повреждения ДНК.

Во второй работе [Puin и др., 2022] мы сопоставляли данные Hi-C с данными RNA-seq и DamID - особым типом эксперимента, позволяющим картировать сайты связывания ДНК- или хроматин-связывающих белков. Мы сравнивали организацию хроматина в модели сперматогониев (SpG) и семенниках, заполненных в основном сперматоцитами (SpCs). Применение разработанных нами методов интеграции Hi-C с другими типами данных позволило показать, что инициация SpCs-специфичной транскрипции коррелирует с отдалением промоторов от ламины и с локальной пространственной изоляцией прилегающих к ним областей. Мы также выявили повышенную частоту контактов между SpCs-специфичными генами в SpCs, что подразумевает их организацию в фабрики транскрипции.

Таким образом, разработанная нами методика сопоставления данных Hi-C с другими омиксными данными позволяет пронаблюдать взаимосвязь между особенностями укладки хроматина и экспрессией генов, уровнем ацетилирования гистонов, связыванием транскрипционных факторов, и другими уровнями молекулярной организации клеток.

Разработка методов анализа липидного и метаболического состава мозга человека

В то время как методы RNA-seq, ChIP-seq, ATAC-seq широко применяются для изучения мозга человека, описание методов, оптимизированных для изучения липидного и метаболического состава мозга человека, в литературе отсутствует. Поэтому мы разработали методики экстракции липидов и масс-спектрометрического анализа, которые обеспечивают одновременное детектирование максимального количества различных липидных классов и индивидуальных веществ в составе мозга человека, а также подходы к биоинформатическому анализу полученных данных. Их применение позволяет создать целостную картину молекулярной организации клеток головного мозга человека.

Экстракцию липидов из образцов ткани проводили согласно процедуре, которая подробно описана в нашей работе [Стеколыщикова и др., 2021] и позволяет детектировать наибольшее количество липидов мозга. Подробное описание процедуры биоинформатического анализа на примере липидомных данных для мозга человека приведено в нашей работе [Smirnov и др., 2021] и может применяться не только для липидомных, но и для метаболомных данных. Пример применения этого подхода к анализу метаболомных данных для ткани мозга представлен в работе [Smirnov и др., 2023] и описан в разделе Результаты и обсуждение диссертационной работы.

Таким образом, нам удалось разработать необходимый набор инструментов и подходов для анализа данных Hi-C и других типов омиксных данных. Пока данные Hi-C для мозга человека не были доступны, инструменты тестировались на иных данных. Однако, мы уделяли большое внимание универсальности разрабатываемых методов, всегда тестируя их на разных наборах данных, чтобы каждый из подходов можно было впоследствии применить к появившимся данным для мозга человека. Важно отметить, что в процессе тестирования мы сделали ряд новых наблюдений об организации хроматина на всех иерархических уровнях - от хромосомных территорий и компартментов до ТАДов и петель, а также о регуляторных аспектах функционирования клеток. Эти наблюдения позволили существенно расширить имеющиеся в мировой литературе фундаментальные представления о регуляторных процессах, структуре и функциях хроматина, универсальных для клеток разных типов.

Результаты и обсуждение

Третья глава посвящена результатам диссертационной работы и их обсуждению. Прежде, чем переходить к изучению архитектуры хроматина в мозге человека, необходимо было исследовать общие принципы организации хроматина, чтобы поместить наблюдаемые особенности упаковки хроматина в клетках головного мозга в контекст общих представлений, универсальных для всех клеточных типов, о разных уровнях архитектуры хроматина - от крупномасштабных (хромосомные территории, компартменты) до самых мелких (ТАДы, хроматиновые петли).

Изучение общих принципов организации хроматина

Хромосомные территории. На формирование и позиционирование хромосомных территорий влияют различные факторы, такие как размер хромосом, плотность генов и транскрипционная активность. Участки хромосом со сходной транскрипционной активностью имеют тенденцию группироваться

вместе, образуя транскрипционные фабрики, которые облегчают скоординированную регуляцию генов.

В работе [Kaufmann и др., 2015] мы детально изучили сеть межхромосомных контактов хроматина, и показали их взаимосвязь с функциональным подобием генов в клетках человека и мыши. Пространственная близость участков оказалась скоррелирована с функциональным подобием генов в обоих организмах, и с ко-экспрессией генов у человека. Также, взаимодействующие участки обогащены активными маркерами хроматина в человеческих клетках. У мыши такого обогащения не наблюдается. Таким образом, мы можем сделать вывод, что меж-хромосомные взаимодействия не консервативны между видами, в отличие от внутри-хромосомных взаимодействий.

В то же время, расположение хромосомных территорий является высококонсервативным и неслучайным. Аберрантная организация хромосомных территорий способствует нестабильности генома за счет образования хромосомных транслокаций и нарушения регуляции механизмов репарации ДНК. К нестабильности генома и нарушению регуляции механизмов репарации ДНК также приводит нокаут SIRT6 или естественное понижение уровня SIRT6 при старении клеток, т.к. SIRT6 является одним из главных факторов, который поддерживает стабильность генома и способствует репарации ДНК. SIRT6 является ядерным, ассоциированным с хроматином белком, который выполняет функцию деацетилирования гистонов и участвует в организации хроматина. В работе [Smirnov и др., 2023] мы определили ещё одну важную функцию деацетилазы SIRT6 - участие в регуляции митохондриальных процессов в головном мозге.

Наш анализ данных ChIP-seq показал, что SIRT6 регулирует работу митохондриальных генов совместно с белком YY1, т.к. их пики имеют неслучайное перекрывание в промотерах митохондриальных генов. При этом известно, что белок YY1 участвует в механизме образования хроматиновых петель (см. ниже), в частности при формировании нейронов. Этот факт позволяет сделать предположение, что регуляторная функция SIRT6 в мозге также может быть опосредована структурой хроматина, либо напрямую через деацетилазную активность SIRT6, либо через взаимодействие с архитектурным белком YY1.

Компартменты хроматина. Компартментализация генома зависит от типа клеток и подвергается ремоделированию во время дифференцировки клеток, а также меняется во время сперматогенеза, как мы показали в работе [Puin и др., 2022]. А именно, наблюдается существенное ослабление компартментализации при переходе от сперматогониев (SpG) к сперматоцитам (SpCs). Учитывая наблюдаемое нами усиление контактов между плечами хромосом в пространстве в SpCs относительно SpG, этот результат указывает на более сильное перемешивание активного и неактивного компартментов хроматина внутри хромосомных территорий в SpCs по сравнению с SpG, что

может быть связано с наличием увеличенных хромосомных территорий на стадии SpCs.

Механизмы, которые управляют образованием компартментов А и В, до конца не изучены, но считается, что они включают взаимодействие между биофизическими свойствами хроматина, эпигенетическими модификациями и активностью транскрипции.

Топологически ассоциированные домены (ТАДы). На основании результатов анализа Hi-C и РНК секвенирования в четырех клеточных линиях *D. melanogaster*, мы предположили, что особые свойства нуклеосом в активном и репрессированном хроматине играют ключевую роль в формировании ТАДов [Uljanov и др., 2016]. В этой работе мы показали, что ТАДы обеднены активными маркерами хроматина и разделены участками активного хроматина с постоянно экспрессирующимися генами. В отличие от предыдущих исследований клеток млекопитающих, мы не наблюдали сильного обогащения сайтами связывания инсуляторных белков на границах ТАДов. Более того, позиции связывания инсуляторных белков предсказывали границы ТАДов хуже, чем маркеры активного хроматина (см. раздел Материалы и методы, рис. 6). Мы предположили, что функциональные особенности хроматина (степень активности, уровень экспрессии генов) определяют его пространственную структуру, а именно положение границ ТАДов (рис. 8А). Поэтому мы предложили механизм формирования ТАДов, основанный на способности нуклеосом в неактивном хроматине притягиваться друг к другу, и на отсутствии такой способности у ацетилированных нуклеосом.

В более позднем исследовании, выполненном для популяции клеток дрозофилы, мы показали, что ТАДы имеют иерархическую структуру: более мелкие активные ТАДы вложены в крупные неактивные [Luzhin и др., 2019] (рис. 8Б). в отдельных клетках дрозофилы. В геноме отдельных клеток дрозофилы, напротив, наблюдаются неиерархические ТАДы, как мы показали в работе [Uljanov и др., 2021]. В целом, иерархическая организация ТАДов в контексте других уровней организации хроматина имеет сложную природу. Сейчас считается, что ТАДы скорее всего образуются по независимому от компартментализации механизму, в соответствии с моделью экструзии хроматиновых петель.

Хроматиновые петли. Хроматиновые петли можно условно разделить на две категории: расположенные на границах ТАДов петли и расположенные внутри ТАДов петли, которые специфически связывают промоторы и энхансеры. Петли хроматина, обнаруженные на границах ТАДов, формируются с помощью модели петлевой экструзии, в которой когезин, кольцеобразный белковый комплекс, окружает ДНК и постепенно выдавливает ее, образуя петлю, ограниченную сайтами CTCF. В дополнение к когезину и CTCF, в образование и поддержание энхансер-промоторных петель могут быть вовлечены дополнительные белки.

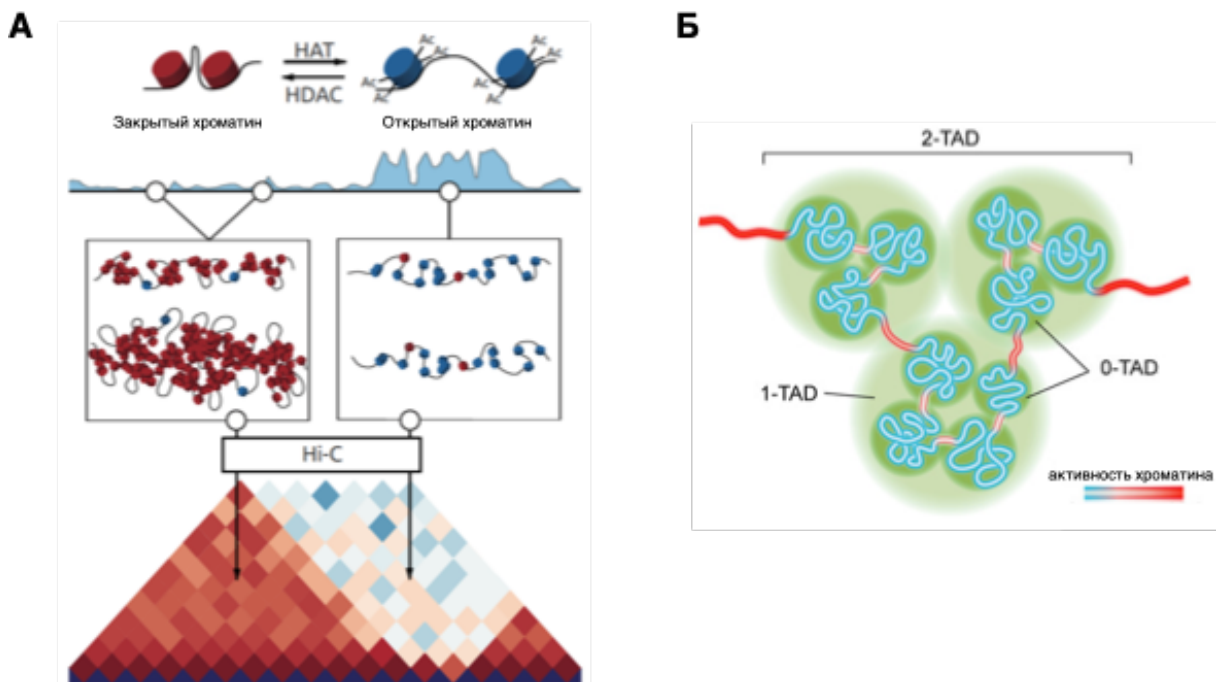


Рис. 8 — Степень активности хроматина определяет его пространственную структуру, а именно положение границ ТАДов. (А) Схема, иллюстрирующая предложенную нами модель сворачивания хроматина в ТАДы в соответствии с ассоциацией нуклеосом. Высокий уровень ацетилирования хроматина в областях генома, содержащих активно транскрибируемые гены, препятствует упаковке хроматина в ТАДы из-за снижения межнуклеосомных взаимодействий. Адаптировано из [Ulianov и др., 2016]. (Б) Схема, иллюстрирующая иерархическую вложенность ТАДов, согласно данным Hi-C для популяций клеток. Адаптировано из [Luzhin и др., 2019].

При формировании нейронов в процессе развития головного мозга человека, главным партнером CTCF в образовании петель является белок YY1, который действует как архитектурный белок, соединяющий петлевые взаимодействия, регулируемые при развитии. В работе [Stein и др., 2021] мы показали, что YY1 играет важную регуляторную роль не только при развитии мозга, но и при его старении, причем как при нормальном старении, так и при патологическом, включая нейродегенеративные заболевания. В этой работе мы выполнили анализ дифференциальной экспрессии (ДЭ) генов между образцами мозга разных возрастов, чтобы проанализировать изменения транскрипции при нормальном старении, а также между образцами с нокаутом SIRT6 (SIRT6-KO), в качестве модели патологического старения (рис. 9А). Поскольку сделать нокаут в мозге человека невозможно, работа выполнялась на образцах мозга мышей. Иерархическая кластеризация образцов продемонстрировала высокую согласованность между биологическими репликациями (рис. 9Б). Интересно, что некоторые гены показали сходные изменения экспрессии при нормальном и патологическом старении (рис. 9В,Г). Мы провели функциональный анализ найденных ДЭ генов и определили транскрипционные факторы, их регулирующие. Белок YY1 оказался в числе наиболее обо-

гашенных транскрипционных факторов. Более того, мы показали, что YY1 образует комплекс с SIRT6 [Stein и др., 2021].

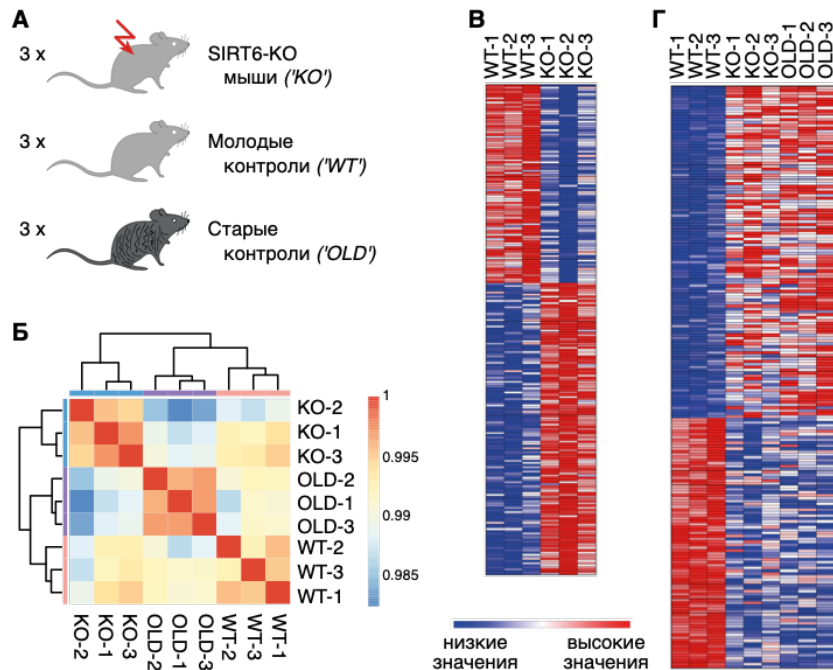


Рис. 9 — Нокаут SIRT6 влияет на экспрессию генов в головном мозге. (А) Дизайн эксперимента. (Б) Иерархическая кластеризация образцов. Цвета показывают коэффициенты корреляции Пирсона. (В) Уровни экспрессии ДЭ генов при сравнении между SIRT6-KO и WT. (Г) Уровни экспрессии ДЭ генов при сравнении объединения SIRT6-KO и старых мышей WT по сравнению с молодыми мышами WT. Адаптировано из [Stein и др., 2021].

Таким образом, YY1 и SIRT6 играют важную регуляторную роль в клетках мозга и могут совместно регулировать экспрессию. Регуляторная функция YY1 точно связана со структурой хроматина, поскольку было показано, что YY1 участвует в формировании хроматиновых петель при нейрогенезе. Кроме того, YY1 может взаимодействовать как с гистондеацетилазами (HDAC1-HDAC3), так и с гистонацетилтрансферазами (p300 и CREB), выступая в качестве контекстно-зависимого регулятора ремоделирования хроматина. Эта универсальность вовлекает YY1 в ключевые механизмы транскрипции внутри клетки, а его деплеция приводит к уменьшению энхансер-промоторных взаимодействий и глобальным изменениям экспрессии генов. Понятно, что деацетилаза SIRT6 также является ключевым игроком в этих процессах, и скорее всего действует через реорганизацию хроматина, однако точный механизм остается неясным и только начал раскрываться в нашей недавней работе [Smirnov и др., 2023].

Изучение организации хроматина в нейронах мозга человека

С использованием оптимизированного нами в работе [Eremenko и др., 2021] протокола Hi-C, мы выполнили эксперименты Hi-C для региона мозга

человека BA22p (левая часть зоны Вернике - латеральный участок височной коры мозга человека, связанный с пониманием речи), с выделением однородной популяции целевых клеток (нейронов) из гетерогенной структуры мозговой ткани путем сортировки ядер на основе окрашивания нейрональным маркером NeuN (NeuN+), а также популяции не-нейрональных клеток (NeuN-) (рис. 10А). Анализ наших карт Hi-C показывает, что хроматин имеет очень специфическую структуру в нейронах, по сравнению с другими клетками мозга (рис. 10Б).

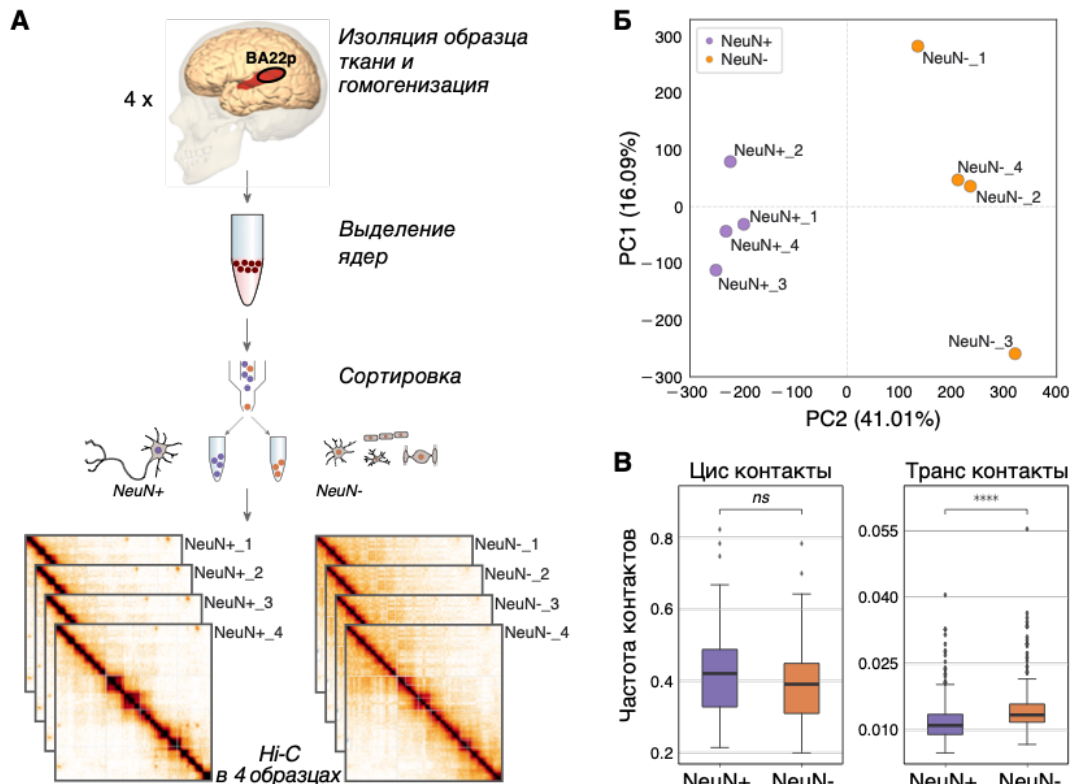


Рис. 10 — Организация хроматина в клетках NeuN+ и NeuN-. (А) Анатомическая локализация анализируемой области мозга, экспериментальная процедура и дизайн исследования. (Б) График анализа главных компонент, основанный на вариации показателя инсуляции (IS) среди всех созданных карт Hi-C. Цвета обозначают клетки NeuN+ и NeuN- здесь и на панели В. (В) Взаимодействия внутри всех хромосом (левая панель) и между всеми парами хромосом (правая панель). Звездочками отмечены Р-значения критерия Вилкоксона: **** - $p < 0,00001$, ns - $p > 0,05$. Адаптировано из [18].

На уровне глобальных различий в организации хроматина, анализ частоты цис- и транс-контактов показывает значительное снижение транс-контактов в нейронах по сравнению с не-нейронами (критерий Вилкоксона, Р-значение $< 10^{-5}$, рис. 10В), что указывает на то, что хромосомные территории в клетках NeuN+ более выражены. Заметно отличающаяся частота контактов на больших геномных расстояниях между популяциями клеток NeuN+ и NeuN- потенциально может быть связана с различной компартиментализацией хроматина в этих клетках. Действительно, как визуальное сравнение, так и анализ главных компонент карт Hi-C подтверждают эту гипотезу.

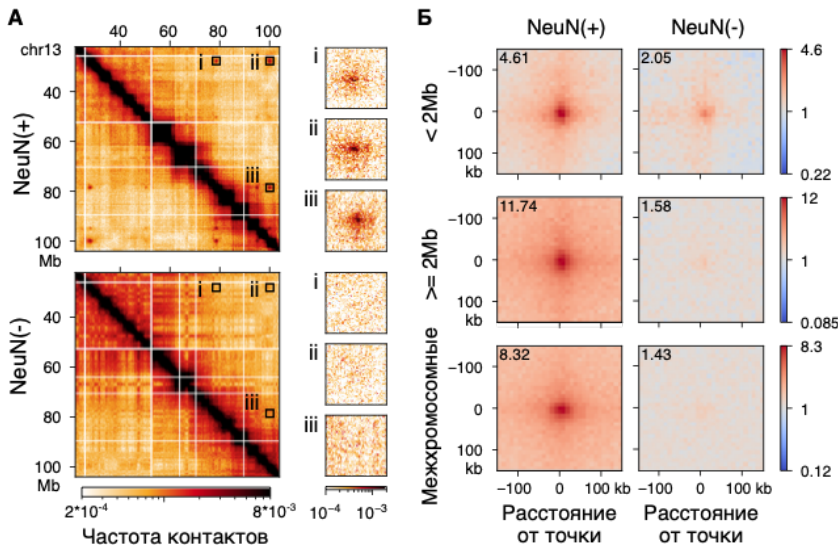


Рис. 11 — Свойства нейрональных точек. (А) Фрагмент карты Hi-C с нейрональными точками, присутствующими в NeuN+ (вверху), но отсутствующими в NeuN- (внизу). (Б) Средний сигнал Hi-C для нейрональных точек (слева) и соответствующих им пар локусов в NeuN- (справа). Адаптировано из [18].

Примечательной особенностью карт NeuN+ Hi-C является наличие ярких точек, расположенных далеко от главной диагонали и, таким образом, представляющих собой сильные контакты на большом расстоянии, 92% (197 из 214) которых перекрываются с пиками H3K27me3 ChIP-seq (далее мы называем эти контакты “нейрональными точками”, рис. 11А). Частота контактов Hi-C в нейрональных точках намного выше по сравнению с частотой контактов тех же пар областей в NeuN- (рис. 11Б). Нейрональные точки образуют сети, включающие до 11 локусов. Эти сети могут включать межхромосомные взаимодействия, которые присутствуют почти исключительно в нейронах.

Из всех генов в локусах нейрональных точек, 80% являются транскрипционными факторами (ТФ), причем они значимо ассоциированы с развитием организма, согласно анализу GO онтологий. Подавляющее большинство этих ТФ высоко экспрессируются в развитии и низко экспрессируются в зрелых NeuN+ и NeuN- клетках. Примечательно, что большинство (119 из 197) из этих ТФ значимо связаны с развитием нервной системы. Поскольку метка H3K27me3 связана с PcG репрессией, мы предполагаем, что белки PcG подавляют гены ТФ в нейрональных точках. Эта гипотеза подтверждается двумя исследованиями, проведенными на мышах [19; 20] и демонстрирующими, что нокауты белков PRC2 или RING1A/B в нейронах мыши приводят к усилению экспрессии многих мышечных ортологов генов из нейрональных точек человека.

Таким образом, 3D организация генома нейронов человека в значительной степени отличается от других типов клеток. Многие биологические процессы, включая экструзию петель, компартиментализацию, дальние взаимодействия, вероятно, работают в нейронах человека по-другому, что способствует правильной экспрессии генов в клетках этого типа. Поэтому изучение функциональной роли организации хроматина в клетках мозга человека невозможно без сопоставления данных Hi-C с данными об экспрессии генов и другими омиксными данными.

Интеграция данных Hi-C с другими омиксными данными в контексте изучения мозга человека

Человек как вид характеризуется высоким уровнем интеллекта, причем потенциал развития сложных когнитивных функций специфичен для людей: детеныши приматов, выросшие в человеческой среде, не развивают сопоставимых когнитивных навыков. Первым шагом на пути к определению механизма развития когнитивных функций в норме и патологии стало детальное описание отличий головного мозга человека от других приматов на молекулярном уровне.

Сравнение транскриптома тканей головного мозга человека с другими приматами. Сравнение головного мозга человека с другими приматами на уровне экспрессии генов представлено в нашей работе [Khrameeva и др., 2020]. Мы проанализировали методом RNA-seq 422 образца мозга человека, шимпанзе, бонобо и макаки (рис. 12А), которые охватывали сразу 33 анатомические области мозга (рис. 12Б). Визуализация варибельности экспрессии среди всех образцов выявила разделение видов в соответствии с их филогенетическим родством (рис. 12В).

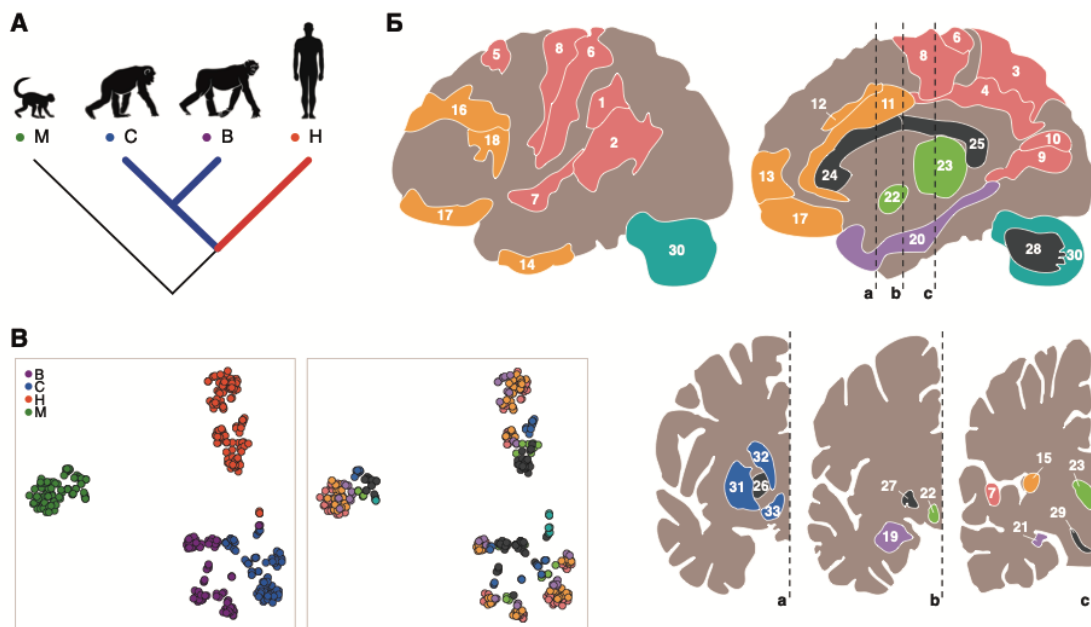


Рис. 12 — Карта транскриптома мозга человека, шимпанзе, бонобо и макаки. (А) Филогенетические взаимоотношения между анализируемыми видами. Цветом выделены ветви, используемые при анализе человек-специфичности экспрессии. (Б) Анатомическая локализация 33 анализируемых областей головного мозга. Цвета обозначают кластеры регионов, определенные нами на основании подобия уровней экспрессии. (В) Графики t-SNE, основанные на варибельности экспрессии для всех 422 проанализированных образцов. Цвета точек обозначают виды (слева) или кластеры регионов (справа).

Адаптировано из [Khrameeva и др., 2020].

Далее, мы оценили человек-специфичность экспрессии в каждой области мозга как отношение числа специфичных для человека различий к коли-

честву специфичных для шимпанзе и бонобо различий. Этот анализ показал избыток специфичных для человека различий в экспрессии в большинстве областей новой коры, а также в гипоталамусе, внутренней капсуле, белом и сером веществе мозжечка.

Для трех из 33 регионов мозга такой же анализ проводился с разрешением в одну клетку методом snRNA-seq. Нейрональные клетки, как правило, показывали меньшую человек-специфичность по сравнению с другими типами клеток. Напротив, астроциты и клетки-предшественники олигодендроцитов стабильно демонстрировали наибольшую человек-специфичность экспрессии, во всех исследованных областях мозга, что указывает на важность глиальных клеток для возникновения новых функций головного мозга у человека.

Эволюция липидного состава тканей мозга человека. Наши исследования экспрессии генов [Khrameeva и др., 2020], некодирующих РНК [Zhang и др., 2016], альтернативного сплайсинга [Pervouchine и др., 2012] позволили понять детали эволюции экспрессии генов на таком уровне детализации, какой ранее был недоступен. Однако глубокое понимание механизмов функционирования мозга невозможно без проведения всестороннего анализа его молекулярной организации. Важный уровень молекулярной организации мозга, которому до наших работ не было уделено внимания, - это уровень небольших молекул, метаболитов и липидов, которые непосредственно вовлечены во все функциональные процессы в мозге и оказались важны для становления *Homo sapiens* как отдельного вида.

Мы сравнили концентрации липидов у человека и более 30 видов млекопитающих в работе [Khrameeva и др., 2018]. Для этого мы измерили содержание липидов в шести тканях 32 видов, представляющих приматов, грызунов и летучих мышей (всего 669 образцов). Исследовались образцы тканей из коры больших полушарий головного мозга, почек, сердца, мышц, печени и мозжечка. Анализ видо-специфичных различий в содержании липидов показал, что линия человека в целом отличается наибольшими отличиями в содержании липидов по сравнению с другими исследованными линиями млекопитающих. При этом оказалось, что у человека липидные изменения локализованы преимущественно в мозге.

Интересно, что найденные нами человек-специфичные липиды преимущественно вовлечены в метаболические пути, связанные с когнитивными заболеваниями, такими как болезнь Альцгеймера, Паркинсона, и другими нарушениями работы нервной системы, что указывает на значимость липидов для когнитивной деятельности - отличительной черты человека как вида. Кроме того, на основании того же самого набора данных мы пришли к выводу, что липидный метаболизм оказывает существенное влияние на максимальную продолжительность жизни того или иного вида в нашей работе [Bozek и др., 2017].

Продолжая исследовать важность липидов для функционирования головного мозга человека в целом и для поддержания человек-специфичных процессов в мозге в частности, в работе [Khrameeva и др., 2014] мы впервые проанализировали липидный состав в образцах префронтальной коры 14 представителей трех человеческих популяций (китайской, европейской и африканской), а также 14 взрослых шимпанзе, и показали, что концентрации липидов значительно изменялись от шимпанзе к человеку в метаболических путях, связанных с катаболизмом липидов (Lipid Catabolic Process, или LCP), у европейцев по сравнению с другими популяциями (рис. 13А). Анализ экспрессии генов по данным RNA-seq, полученным в тех же образцах человека и шимпанзе, подтвердил наши наблюдения на уровне липидома: гены из пути LCP показали максимальное расхождение в экспрессии с шимпанзе у европейцев, промежуточное у азиатов и отсутствующее у африканцев (рис. 13Б). Таким образом, липидные отличия, обнаруженные у европейцев, по-видимому, связаны с изменениями уровней экспрессии соответствующих метаболических ферментов. В поисках возможных причин наблюдаемых отличий, мы обратились к анализу геномных последовательностей генов LCP.

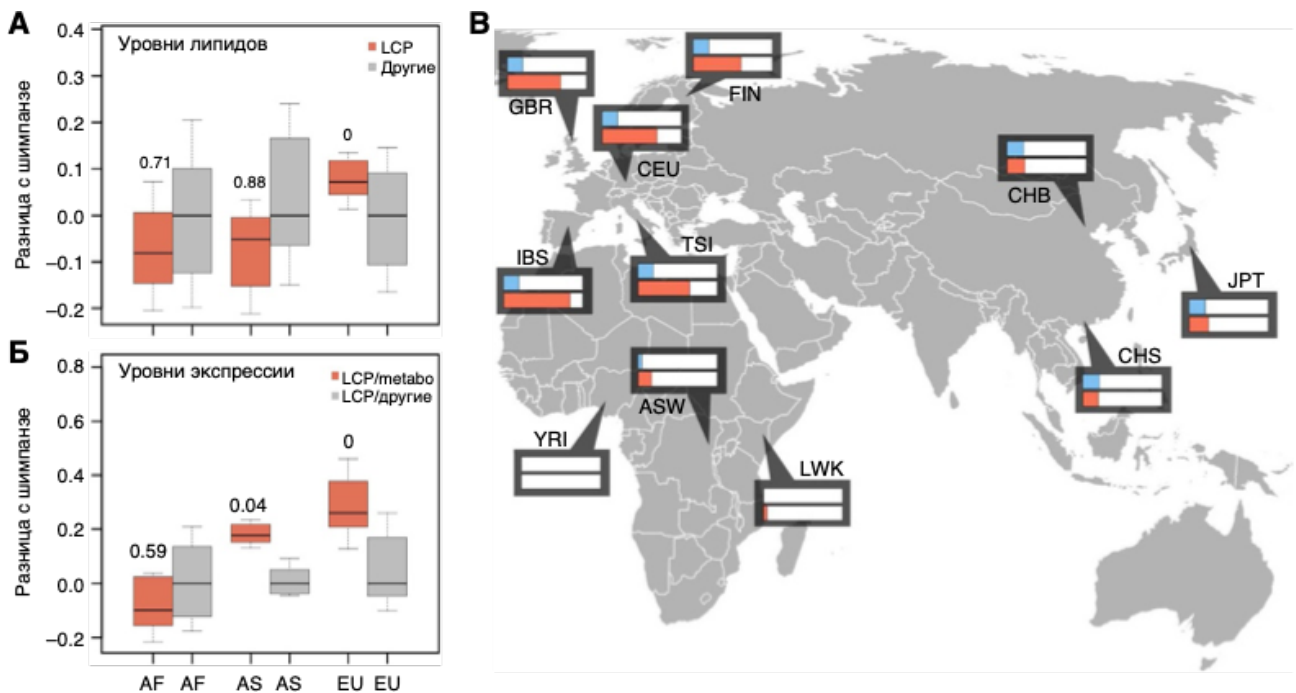


Рис. 13 — Сравнение уровней экспрессии и содержания липидов в префронтальной коре человека и шимпанзе. (А) Отличия содержания липидов между популяциями шимпанзе и людей африканского (AF), азиатского (AS) и европейского (EU) происхождения для участников пути катаболизма липидов (LCP, красный) и метаболитов в других метаболических путях (серый). (Б) Отличия экспрессии генов между популяциями шимпанзе и человека для генов LCP, показанными на панели А (красный) и остальных генов из категории LCP (серый). Цифры на панелях А и Б показывают Р-значения (пермутационный тест). (В) Пропорции NLS в современных популяциях человека. Синий - по всему геному; красный - в генах LCP. Максимальная длина полосы соответствует 30%. Адаптировано из [Khrameeva и др., 2014].

Поскольку уже было известно, что современные европейцы унаследовали от неандертальцев несколько процентов генома, мы хотели проверить гипотезу о том, что наблюдаемые изменения липидного состава у европейцев могут быть связаны с событием скрещивания человека с неандертальцами. В среднем по всему геному, частота неандертальских вариантов (Neanderthal-like sites, или NLS) составила $6,1 \pm 0,2\%$ у современных людей европейского и азиатского происхождения. Более того, частота NLS оказалась выше в три раза в генах LCP, причем только у европейцев ($20,8 \pm 2,6\%$), но не у азиатов ($6,7 \pm 0,7\%$), по сравнению со средним по всему геному значением (рис. 13В).

Избыток NLS в генах LCP в геномах современных европейцев может быть обусловлен быстрым распространением неандертальских аллелей у европейских предков из-за адаптивного значения LCP генов. В частности, можно выдвинуть гипотезу, что со временем неандертальцы приобрели изменения в катаболизме липидов, которые были полезны для выживания в условиях окружающей среды доисторической Европы. Эти адаптивные варианты, возможно, распространились с высокой частотой в результате положительного отбора, поскольку мы обнаружили значимое обогащение признаков положительного отбора в генах LCP у современных европейцев, но не у азиатов или африканцев [Khrameeva и др., 2014].

Нарушения эволюционно новых метаболических путей в мозге человека. Поскольку когнитивные функции, нарушенные при психических расстройствах, являются одними из ключевых поведенческих способностей, которые отделяют человека от других видов, психические расстройства могут представлять собой нарушение эволюционно-новых, специфичных для человека механизмов работы мозга. Наша работа [Kurochkin и др., 2019] была направлена на подтверждение этой гипотезы на уровне метаболитов. Мы изучили префронтальную кору головного мозга 32 пациентов с расстройствами аутистического спектра (РАС) разного возраста (2-60 лет) и 40 соответствующих им контролей (0-61 лет), где удалось обнаружить 202 метаболита с изменениями содержания при РАС - около 15% от общего количества измеренных метаболитов.

Гены, связанные со специфичными для человека метаболитами (по сравнению с шимпанзе и макаками), были значимо перепредставлены в 27 путях KEGG (гипергеометрический тест, P-значение $< 0,05$ с поправкой на множественное тестирование). Примечательно, что эти пути значимо перекрывались с путями, обогащенными различиями в интенсивности метаболитов, связанными с РАС, и включали метаболизм пуринов и пиримидинов (критерий Фишера, P-значение $< 0,01$), к которому относится фермент аденилосукцината-лиаза (ADSL) - один из немногих белков, несущих аминокислотную замену, которая присутствует сегодня у всех людей, но отсутствует у неандертальцев. Функциональная роль этой специфичной для человека мутации была показана позже в другой нашей работе [Stepanova и др., 2021] на мышинных моделях,



Рис. 14 — Перекрытие между энхансерами, связанными с человек-специфичными генами, и активными в мозге цис-регуляторными элементами, расположенными в HAR [21]. Гистограмма представляет собой распределение случайных значений перекрывания. Красной пунктирной линией отмечено фактическое перекрывание. Адаптировано из [Khrameeva и др., 2020].

подтвердив гипотезу о том, что изменение метаболизма пуринов произошло у человека после его отделения от предка, общего с неандертальцами и денисовцами.

Роль организации хроматина в регуляции специфичной для человека экспрессии в мозге. Чтобы понять возможные причины наблюдаемых нами эволюционных изменений на человеческой ветке, мы изучили контакты энхансеров и промоторов генов с уникальной для определенного региона мозга человека экспрессией, на основе данных о трехмерной организации хроматина. Мы отобрали 1271 уникальных генов, показывающих специфические для человека уровни экспрессии на основе данных с разрешением в одну клетку (см. подробное описание процедуры отбора в нашей работе [Khrameeva и др., 2020]). Используя базу данных GeneHancer, мы сформировали список энхансеров, связанных с этими генами. Важно отметить, что соответствие между энхансерами и генами в базе GeneHancer установлено на основе данных о трехмерной организации хроматина, поскольку, как подробно обсуждалось выше, энхансер-промоторные взаимодействия поддерживаются хроматиновыми петлями, которые специфически связывают промоторы и энхансеры при участии ряда вспомогательных белков, таких как YY1.

Затем мы подсчитали энхансеры, соответствующие активным в мозгу цис-регуляторным элементам, расположенным в областях с ускоренной эволюцией в геноме человека (Human Accelerated Regions, или HAR). Мы использовали 284 предсказанных HAR из работы [21]. Гены, демонстрирующие специфическую для человека экспрессию действительно располагались в непосредственной близости от этих энхансеров (перекрывающихся с активными в мозгу цис-регуляторными элементами, расположенными в HAR) чаще, чем ожидалось случайно (98 из 248 HAR, рис. 14). Чтобы оценить значимость этого совпадения, мы выполнили пермутационный тест: 1000 раз сделали случайные подвыборки размером 1271 ген из 9138 генов, экспрессированных в головном мозге (P -значение $< 0,005$, рис. 14).

Для анализа сайтов связывания транскрипционных факторов (ТФ) в этих 98 энхансерах, перекрывающихся с HAR, мы использовали распреде-

ление сайтов связывания ТФ из базы данных GeneHancer, составленное на основе анализа данных ChIP-seq. 35 из 42 ТФ, демонстрирующих специфическую для человека экспрессию в наших данных, имеют сайты связывания в этих 98 энхансерах, перекрывающихся с НАР. Более того, 13 из этих 35 ТФ связываются с 98 энхансерами, перекрывающимися с НАР, чаще, чем ожидалось случайно.

Данный анализ, основанный на данных о трехмерной организации хроматина, придавал большую значимость нашему исследованию, установив возможную причину наблюдаемых нами эволюционных изменений на человеческой ветке.

Заключение

В заключении приводятся основные **выводы** работы, которые состоят в следующем:

1. Разработана методология биоинформатического анализа данных Hi-C для образцов мозга человека: построены карты контактов, проведен анализ хроматиновых компартментов, ТАДов и петель.

2. Решена проблема разной экспериментальной представленности геномных регионов в данных Hi-C.

3. Разработана методика выявления биологического сигнала в данных Hi-C с высоким уровнем технической вариабельности.

4. Оптимизирован метод аннотации ТАДов на основе обогащения эпигенетическими метками.

5. Методы машинного обучения адаптированы и применены для исследования организации хроматина.

6. С помощью разработанных подходов к анализу данных Hi-C, описаны особенности трехмерной организации генома в нейронах и других клетках мозга человека, а также установлена функциональная роль дальних взаимодействий, которые опосредованы PcG белками и присутствуют только в нейронах.

7. Раскрыты общие принципы организации хроматина, такие как иерархичность ТАДов, в том числе в одиночных клетках, роль ТАДов в регуляции транскрипции, роль ядерной ламины, и др., что позволяет поместить особенности организации хроматина нейронов в контекст общих представлений о структуре хроматина.

8. Разработаны методы анализа данных RNA-seq, ATAC-seq, ChIP-seq и др. для решения нестандартных задач в контексте изучения мозга человека, таких как анализ биаллельной транскрипции, механизмы регуляции альтернативного сплайсинга (данные iCLIP), анализ некодирующих РНК, разделение сигнала от тканей человека и мыши в ксенотрансплантатах (PDX) для

изучения онкозаболеваний, вычислительное восстановление видовой принадлежности отдельных клеток (данные snRNA-seq).

9. При помощи этих методов, а также разработанных методов анализа липидного и метаболического состава мозга человека, установлена роль липидов и конкретных метаболитов в функционировании клеток головного мозга человека в норме и при заболеваниях, в частности при психических расстройствах.

10. Оценена роль неандертальских вариантов в качестве возможной причины липидных отличий, специфичных для европейской популяции, и оценена роль неандертальских вариантов в других популяциях, в частности в популяциях кетов и папуасов.

11. Разработаны методы интеграции данных Hi-C с другими омиксными данными: RNA-seq, ATAC-seq, ChIP-seq и др., при помощи которых функционирование клеток мозга описано комплексно, что позволило установить регуляторную роль деацетилазы SIRT6 и архитектурного белка YY1, участвующего в образовании хроматиновых петель.

12. Установлено, что организация хроматина участвует в регуляции экспрессии генов в мозге человека, в особенности в участках генома, характеризующихся ускоренной эволюцией на человеческой ветке.

Таким образом, в данной работе установлена роль трехмерной организации генома в регуляции функционирования человеческого мозга, с учетом гетерогенности его клеточного состава и в комплексе с другими уровнями молекулярной организации клеток: экспрессией генов, эпигенетическими модификациями, связыванием транскрипционных факторов, открытостью хроматина и метаболизмом клеток в целом.

Публикации автора по теме диссертации

1. Kobets VA, Ulianov SV, Galitsyna AA, Doronin SA, Mikhaleva EA, Gelfand MS, Shevelyov YY, Razin SV, **Khrameeva EE**. HiConfidence: a novel approach uncovering the biological signal in Hi-C data affected by technical biases. *Brief Bioinform.* 2023; 24(2):bbad044. doi: 10.1093/bib/bbad044.

2. Smirnov D, Eremenko E, Stein D, Kaluski S, Jasinska W, Cosentino C, Martinez-Pastor B, Brotman Y, Mostoslavsky R, **Khrameeva E[#]**, Toiber D[#]. SIRT6 is a key regulator of mitochondrial function in the brain. *Cell Death Dis.* 2023; 14(1):35. doi: 10.1038/s41419-022-05542-w.

3. Shcherbinina E, Abakumova T, Bobrovskiy D, Kurochkin I, Deinichenko K, Stekolshchikova E, Anikanov N, Ziganshin R, Melnikov P, **Khrameeva E**, Logacheva M, Zatsepin T, Sergeeva O. Murine Falcor/LL35 lncRNA Contributes to Glucose and Lipid Metabolism In Vitro and In Vivo. *Biomedicines.* 2022; 10(6):1397. doi: 10.3390/biomedicines10061397.

4. Tkachev A, Stekolshchikova E, Bobrovskiy DM, Anikanov N, Ogurtsova P, Park DI, Horn AKE, Petrova D, **Khrameeva E[#]**, Golub MS[#], Turck CW[#], Khaitovich P[#]. Long-Term

Fluoxetine Administration Causes Substantial Lipidome Alteration of the Juvenile Macaque Brain. *Int J Mol Sci.* 2021; 22(15):8089. doi: 10.3390/ijms22158089.

5. Ilyin AA, Kononkova AD, Golova AV, Shloma VV, Olenkina OM, Nenasheva VV, Abramov YA, Kotov AA, Maksimov DA, Laktionov PP, Pindyurin AV, Galitsyna AA, Ulianov SV, **Khrameeva EE**[#], Gelfand MS, Belyakin SN, Razin SV, Shevelyov YY[#]. Comparison of genome architecture at two stages of male germline cell differentiation in *Drosophila*. *Nucleic Acids Res.* 2022; 50(6):3203-3225. doi: 10.1093/nar/gkac109.

6. Stepanova V, Moczulska KE, Vacano GN, Kurochkin I, Ju X, Riesenberger S, Macak D, Maricic T, Dombrowski L, Schörnig M, Anastassiadis K, Baker O, Naumann R, **Khrameeva E**, Vanushkina A, Stekolshchikova E, Egorova A, Tkachev A, Mazzarino R, Duval N, Zubkov D, Giavalisco P, Wilkinson TG, Patterson D, Khaitovich P, Pääbo S. Reduced purine biosynthesis in humans after their divergence from Neandertals. *Elife.* 2021; 10:e58741. doi: 10.7554/eLife.58741.

7. Smirnov D, Mazin P, Osetrova M, Stekolshchikova E, **Khrameeva E**. The Hitchhiker's Guide to Untargeted Lipidomics Analysis: Practical Guidelines. *Metabolites.* 2021; 11(11):713. doi: 10.3390/metabo11110713.

8. Е.А. Стекольников, А.И. Ткачев, М.С. Осетрова, Д.Н. Смирнов, О.И. Ефимова, **Е.Е. Храмеева**. Применение методов масс-спектрометрии к анализу липидного состава головного мозга человека. 2021; Т. 37, № 5, С. 80–87. doi: 10.21519/0234-2758-2021-37-5-80-87.

9. Eremenko E, Golova A, Stein D, Einav M, **Khrameeva E**, Toiber D. FACS-based isolation of fixed mouse neuronal nuclei for ATAC-seq and Hi-C. *STAR Protoc.* 2021; 2(3):100643. doi: 10.1016/j.xpro.2021.100643.

10. Mazor G, Smirnov D, Ben David H, **Khrameeva E**, Toiber D, Rotblat B. TP73-AS1 is induced by YY1 during TMZ treatment and highly expressed in the aging brain. *Aging (Albany NY).* 2021; 13(11):14843-14861. doi: 10.18632/aging.203182.

11. Badarni M, Prasad M, Golden A, Bhattacharya B, Levin L, Yegodayev KM, Dimitstein O, Joshua BZ, Cohen L, **Khrameeva E**, Kong D, Porgador A, Braiman A, Grandis JR, Rotblat B, Elkabets M. IGF2 Mediates Resistance to Isoform Selective-Inhibitors of the PI3K in HPV Positive Head and Neck Cancer. *Cancers (Basel).* 2021; 13(9):2250. doi: 10.3390/cancers13092250.

12. Rozenwald MB, Galitsyna AA, Sapunov GV, **Khrameeva EE**, Gelfand MS. A machine learning framework for the prediction of chromatin folding in *Drosophila* using epigenetic features. *PeerJ Comput Sci.* 2020; 6:e307. doi: 10.7717/peerj-cs.307.

13. Stein D, Mizrahi A, Golova A, Saretzky A, Venzor AG, Slobodnik Z, Kaluski S, Einav M, **Khrameeva E**, Toiber D. Aging and pathological aging signatures of the brain: through the focusing lens of SIRT6. *Aging (Albany NY).* 2021; 13(5):6420-6441. doi: 10.18632/aging.202755.

14. Ulianov SV, Zakharova VV, Galitsyna AA, Kos PI, Polovnikov KE, Flyamer IM, Mikhaleva EA, **Khrameeva EE**, Germini D, Logacheva MD, Gavrilov AA, Gorsky AS, Nechaev SK, Gelfand MS, Vassetzky YS, Chertovich AV, Shevelyov YY, Razin SV. Order and stochasticity in the folding of individual *Drosophila* genomes. *Nat Commun.* 2021; 12(1):41. doi: 10.1038/s41467-020-20292-z.

15. Samborskaia MD, Galitsyna A, Pletenev I, Trofimova A, Mironov AA, Gelfand MS, **Khrameeva EE**. Cumulative contact frequency of a chromatin region is an intrinsic property linked to its function. *PeerJ.* 2020; 8:e9566. doi: 10.7717/peerj.9566.

16. **Khrameeva E**^{*}, Kurochkin I^{*}, Han D^{*}, Guijarro P, Kanton S, Santel M, Qian Z, Rong S, Mazin P, Sabirov M, Bulat M, Efimova O, Tkachev A, Guo S, Sherwood CC, Camp JG, Pääbo

S, Treutlein B, Khaitovich P. Single-cell-resolution transcriptome map of human, chimpanzee, bonobo, and macaque brains. *Genome Res.* 2020; 30(5):776-789. doi: 10.1101/gr.256958.119.

17. Yegodayev KM, Novoplansky O, Golden A, Prasad M, Levin L, Jagadeeshan S, Zorea J, Dimitstein O, Joshua BZ, Cohen L, **Khrameeva E**[#], Elkabets M[#]. TGF-Beta-Activated Cancer-Associated Fibroblasts Limit Cetuximab Efficacy in Preclinical Models of Head and Neck Cancer. *Cancers (Basel).* 2020; 12(2):339. doi: 10.3390/cancers12020339.

18. Tkachev A, Stepanova V, Zhang L, **Khrameeva E**, Zubkov D, Giavalisco P, Khaitovich P. Differences in lipidome and metabolome organization of prefrontal cortex among human populations. *Sci Rep.* 2019; 9(1):18348. doi: 10.1038/s41598-019-53762-6.

19. Kurochkin I*, **Khrameeva E***, Tkachev A, Stepanova V, Vanyushkina A, Stekolshchikova E, Li Q, Zubkov D, Shichkova P, Halene T, Willmitzer L, Giavalisco P, Akbarian S, Khaitovich P. Metabolome signature of autism in the human prefrontal cortex. *Commun Biol.* 2019; 2:234. doi: 10.1038/s42003-019-0485-4.

20. Luzhin AV, Flyamer IM, **Khrameeva EE**, Ulianov SV, Razin SV, Gavrilov AA. Quantitative differences in TAD border strength underly the TAD hierarchy in Drosophila chromosomes. *J Cell Biochem.* 2019; 120(3):4494-4503. doi: 10.1002/jcb.27737.

21. Ulianov SV*, Doronin SA*, **Khrameeva EE***, Kos PI*, Luzhin AV, Starikov SS, Galitsyna AA, Nenasheva VV, Ilyin AA, Flyamer IM, Mikhaleva EA, Logacheva MD, Gelfand MS, Chertovich AV, Gavrilov AA, Razin SV, Shevelyov YY. Nuclear lamina integrity is required for proper spatial organization of chromatin in Drosophila. *Nat Commun.* 2019; 10(1):1176. doi: 10.1038/s41467-019-09185-y.

22. **Khrameeva E***, Kurochkin I*, Bozek K, Giavalisco P, Khaitovich P. Lipidome Evolution in Mammalian Tissues. *Mol Biol Evol.* 2018; 35(8):1947-1957. doi: 10.1093/molbev/msy097.

23. Akkuratov EE, Gelfand MS, **Khrameeva EE**. Neanderthal and Denisovan ancestry in Papuans: A functional study. *J Bioinform Comput Biol.* 2018; 16(2):1840011. doi: 10.1142/S0219720018400115.

24. Ulianov SV, Galitsyna AA, Flyamer IM, Golov AK, **Khrameeva EE**, Imakaev MV, Abdennur NA, Gelfand MS, Gavrilov AA, Razin SV. Activation of the alpha-globin gene expression correlates with dramatic upregulation of nearby non-globin genes and changes in local and large-scale chromatin spatial structure. *Epigenetics and Chromatin.* 2017; 10(1):35. doi: 10.1186/s13072-017-0142-4.

25. Bozek K, **Khrameeva EE**, Reznick J, Omerbašić D, Bennett NC, Lewin GR, Azpurua J, Gorbunova V, Seluanov A, Regnard P, Wanert F, Marchal J, Pifferi F, Aujard F, Liu Z, Shi P, Pääbo S, Schroeder F, Willmitzer L, Giavalisco P, Khaitovich P. Lipidome determinants of maximal lifespan in mammals. *Sci Rep.* 2017; 7(1):5. doi: 10.1038/s41598-017-00037-7.

26. Gavrilov AA, Shevelyov YY, Ulianov SV, **Khrameeva EE**, Kos P, Chertovich A, Razin SV. Unraveling the mechanisms of chromatin fibril packaging. *Nucleus.* 2016; 7(3):319-24. doi: 10.1080/19491034.2016.1190896.

27. **Khrameeva EE**, Fudenberg G, Gelfand MS, Mirny LA. History of chromosome rearrangements reflects the spatial organization of yeast chromosomes. *J Bioinform Comput Biol.* 2016; 1641002. doi: 10.1142/S021972001641002X.

28. Zhang B, Han D, Korostelev Y, Yan Z, Shao N, **Khrameeva E**, Velichkovsky BM, Chen YP, Gelfand MS, Khaitovich P. Changes in snoRNA and snRNA Abundance in the Human, Chimpanzee, Macaque, and Mouse Brain. *Genome Biol Evol.* 2016; 8(3):840-50. doi: 10.1093/gbe/evw038.

29. Flegontov P, Changmai P, Zidkova A, Logacheva MD, Altınışık NE, Flegontova O, Gelfand MS, Gerasimov ES, **Khrameeva EE**, Konovalova OP, Neretina T, Nikolsky YV, Starostin G, Stepanova VV, Travinsky IV, Tříška M, Tříška P, Tatarinova TV. Genomic study of

the Ket: a Paleo-Eskimo-related ethnic group with significant ancient North Eurasian ancestry. *Sci Rep.* 2016; 6:20768. doi: 10.1038/srep20768.

30. Ulianov SV*, **Khrameeva EE***, Gavrilov AA, Flyamer IM, Kos P, Mikhaleva EA, Penin AA, Logacheva MD, Imakaev MV, Chertovich A, Gelfand MS, Shevelyov YY, Razin SV. Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Res.* 2016; 26(1):70-84. doi: 10.1101/gr.196006.115.

31. Kaufmann S, Fuchs C, Gonik M, **Khrameeva EE**, Mironov AA, Frishman D. Inter-chromosomal contact networks provide insights into Mammalian chromatin organization. *PLoS One.* 2015; 10(5):e0126125. doi: 10.1371/journal.pone.0126125.

32. Zhenilo S, **Khrameeva E**, Tsygankova S, Zhigalova N, Mazur A, Prokhortchouk E. Individual genome sequencing identified a novel enhancer element in exon 7 of the CSFR1 gene by shift of expressed allele ratios. *Gene.* 2015; 566(2):223-8. doi: 10.1016/j.gene.2015.04.053.

33. Yang J, Hung LH, Lich T, Kostin S, Looso M, **Khrameeva E**, Bindereif A, Schneider A, Braun T. RBM24 is a major regulator of muscle-specific alternative splicing. *Dev Cell.* 2014; 31(1):87-99. doi: 10.1016/j.devcel.2014.08.025.

34. **Khrameeva EE**, Bozek K, He L, Yan Z, Jiang X, Wei Y, Tang K, Gelfand MS, Prufer K, Kelso J, Paabo S, Giavalisco P, Lachmann M, Khaitovich P. Neanderthal ancestry drives evolution of lipid catabolism in contemporary Europeans. *Nat Commun.* 2014; 5:3584. doi: 10.1038/ncomms4584.

35. Rossbach O, Hung LH, **Khrameeva E**, Schreiner S, König J, Curk T, Zupan B, Ule J, Gelfand MS, Bindereif A. Crosslinking-immunoprecipitation (iCLIP) analysis reveals global regulatory roles of hnRNP L. *RNA Biol.* 2014; 11(2):146-55. doi: 10.4161/rna.27991.

36. Rösel-Hillgärtner TD, Hung LH, **Khrameeva E**, Le Querrec P, Gelfand MS, Bindereif A. A novel intra-U1 snRNP cross-regulation mechanism: alternative splicing switch links U1C and U1-70K expression. *PLoS Genet.* 2013; 9(10):e1003856. doi: 10.1371/journal.pgen.1003856.

37. **Khrameeva EE**, Gelfand MS. Biases in read coverage demonstrated by interlaboratory and interplatform comparison of 117 mRNA and genome sequencing experiments. *BMC Bioinformatics.* 2012; 13 Suppl 6:S4. doi: 10.1186/1471-2105-13-S6-S4.

38. **Khrameeva EE**, Mironov AA, Fedonin GG, Khaitovich P, Gelfand MS. Spatial proximity and similarity of the epigenetic state of genome domains. *PLoS One.* 2012; 7(4):e33947. doi: 10.1371/journal.pone.0033947.

39. Pervouchine DD, **Khrameeva EE**, Pichugina MY, Nikolaienko OV, Gelfand MS, Rubtsov PM, Mironov AA. Evidence for widespread association of mammalian splicing and conserved long-range RNA structures. *RNA.* 2012; 18(1):1-15. doi: 10.1261/rna.029249.111.

* - первые авторы с равным вкладом

- авторы, ответственные за переписку, с равным вкладом

Список литературы

1. Recent evidence that TADs and chromatin loops are dynamic structures / A. S. Hansen [и др.] // *Nucleus.* — 2018. — 31 дек. — т. 9, № 1. — с. 20—32.
2. *Cremer T., Cremer C.* Chromosome territories, nuclear architecture and gene regulation in mammalian cells // *Nature Reviews Genetics.* — 2001. — 1 апр. — т. 2, № 4. — с. 292—301.
3. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome / E. Lieberman-Aiden [и др.] // *Science.* — 2009. — 9 окт. — т. 326, № 5950. — с. 289—293.

4. *Bonev B., Cavalli G.* Organization and function of the 3D genome // *Nature Reviews Genetics*. — 2016. — нояб. — т. 17, № 11. — с. 661–678.
5. Topological domains in mammalian genomes identified by analysis of chromatin interactions / J. R. Dixon [и др.] // *Nature*. — 2012. — май. — т. 485, № 7398. — с. 376–380.
6. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes / A. L. Sanborn [и др.] // *Proceedings of the National Academy of Sciences*. — 2015. — 24 нояб. — т. 112, № 47.
7. *Rowley M. J., Corces V. G.* Organizational principles of 3D genome architecture // *Nature Reviews Genetics*. — 2018. — дек. — т. 19, № 12. — с. 789–800.
8. Chromatin architecture reorganization during stem cell differentiation / J. R. Dixon [и др.] // *Nature*. — 2015. — 19 февр. — т. 518, № 7539. — с. 331–336.
9. Widespread rearrangement of 3D chromatin organization underlies polycomb-mediated stress-induced silencing / L. Li [и др.] // *Molecular cell*. — 2015. — т. 58, № 2. — с. 216–231.
10. Evolutionarily conserved principles predict 3D chromatin organization / M. J. Rowley [и др.] // *Molecular cell*. — 2017. — т. 67, № 5. — с. 837–852.
11. From compartments to gene loops: Functions of the 3D genome in the human brain : preprint / S. Rahman [и др.] ; *Genomics*. — 14.10.2021.
12. Neuronal and glial 3D chromatin architecture informs the cellular etiology of brain disorders / B. Hu [и др.] // *Nature Communications*. — 2021. — 25 июня. — т. 12, № 1. — с. 3968.
13. Molecular and cellular reorganization of neural circuits in the human lineage / A. M. Sousa [и др.] // *Science*. — 2017. — т. 358, № 6366. — с. 1027–1032.
14. Human-specific features of spatial gene expression and regulation in eight brain regions / C. Xu [и др.] // *Genome research*. — 2018. — т. 28, № 8. — с. 1097–1110.
15. A chromosome folding intermediate at the condensin-to-cohesin transition during telophase / K. Abramo [и др.] // *Nature cell biology*. — 2019. — т. 21, № 11. — с. 1393–1402.
16. optimalTAD: annotation of topologically associating domains based on chromatin marks enrichment / D. Smirnov [и др.] // *bioRxiv*. — 2023. — с. 2023–03.
17. Independence of chromatin conformation and gene regulation during *Drosophila* dorsoventral patterning / E. Ing-Simmons [и др.] // *Nat Genet*. — 2021. — апр. — т. 53, № 4. — с. 487–499.
18. Extensive long-range polycomb interactions and weak compartmentalization are hallmarks of human neuronal 3D genome / I. A. Pletenev [и др.] // *bioRxiv*. — 2023. — с. 2023–08.
19. PRC1 sustains the integrity of neural fate in the absence of PRC2 function / A. Sawai [и др.] // *Elife*. — 2022. — т. 11. — e72769.
20. Polycomb repressive complex 2 (PRC2) silences genes responsible for neurodegeneration / M. von Schimmelmann [и др.] // *Nature neuroscience*. — 2016. — т. 19, № 10. — с. 1321–1330.
21. Epigenomic annotation of gene regulatory alterations during evolution of the primate brain / M. W. Vermunt [и др.] // *Nature neuroscience*. — 2016. — т. 19, № 3. — с. 494–503.